32. Rynearson, E.K. (ed.) (2006). *Violent death: resilience and intervention beyond the crisis*. Routledge Psychosocial Stress Series, Taylor & Francis, New York.
33. Vanderwerker, L.C. and Prigerson, H.G. (2004). Social support and technological connectedness as protective factors in bereavement. *Journal of Loss and Trauma*, **9**, 45–57.
34. Parkes, C.M., Laungani, P., and Young, B. (1997). *Death and bereavement across cultures*. Routledge, London.
35. Prigerson, H., Horowitz, M.J., Selby, C.J., *et al*. (manuscript submitted for publication). Field trial of consensus criteria for prolonged grief disorder proposed for DSM-V. *Archives of General Psychiatry*.
36. Stroebe, M., Schut, H., and Stroebe, W. (2007). The physical and mental health consequences of bereavement: a review. Seminar for *Lancet,* **8**, 1960–73.
37. Neimeyer, R.A. and Hogan, N.S. (2001). Quantitative or qualitative? Measurement issues in the study of grief. In *Handbook of bereavement research: consequences, coping, and care* (eds. M.S. Strobe, R.O. Hansson, W. Stroebe, and H. Schut), pp. 89–118. American Psychological Association, Washington, DC.
38. National Center for PTSD Psychological First Aid Manual http://www.ncptsd.va.gov/ncmain/ncdocs/manuals/nc_manual_psyfirstaid.html
39. Raphael, B. (1977). Preventive intervention with the recently bereaved. *Archives of General Psychiatry*, **34**, 1450–4.
40. Vachon, M.L.S., Lyall, W.A.L., Rogers, J., *et al*. (1980). A controlled study of self-help interventions for widows. *The American Journal of Psychiatry*, **137**, 1380–4.
41. Shear, K., Frank, E., Houck, P.R., *et al*. (2005). Treatment of complicated grief: a randomized controlled trial. *The Journal of the American Medical Association*, **293**, 2601–8.
42. Shear, K. and Frank, E. (2006). Treatment of complicated grief: integrating cognitive–behavioral methods with other treatment approaches. In *Cognitive behavioural therapies for trauma* (eds. V. Follette and J. Ruzek), pp. 290–320. Guilford Press, New York.
43. Raphael, B., Dunsmore, J., and Wooding, S. (2004). Early mental health interventions for traumatic loss in adults. In *Early intervention for trauma and traumatic loss* (ed. B. Litz), pp. 147–78. Guilford Press, New York.
44. Wagner, B., Knaevelsrud, C., and Maercker, A. (2006). Internet-based cognitive–behaviorial therapy (INTERAPY) for complicated grief: a randomized controlled trial. *Death Studies*, **30**, 429–53.
45. Rando, T.A. (1993). *The treatment of complicated mourning*. Research Press, Champaign, IL.
46. Murray, J.A., Terry, D.J., Vance, J.C., *et al*. (2000). Effects of a program of intervention on parental distress following infant death. *Death Studies*, **24**, 275–305.
47. Worden, J.W. (1991). *Grief counselling and grief therapy*. Springer Publishing Company, New York.
48. Boelen, P.A. (2005). *Complicated grief: assessment, theory, and treatment*. Ipskamp, Enscede/Amsterdam.
49. Raphael, B., Dobson, M., and Minkov, C. (2001). Psychotherapeutic and pharmacological intervention for bereaved people chapter in *Handbook of bereavement research: consequences, coping, and care* (eds. M.S. Stroebe, R.O. Hansson, W. Stroebe, and H. Schut). American Psychological Association, Washington, DC.
50. Schut, H., Stroebe, M.S., Van Den Bout, J., *et al*. (2001). The efficacy of bereavement interventions: determining who benefits. In *Handbook of bereavement research: consequences, coping, and care* (eds. M.S. Stroebe, R.O. Hansson, W. Stroebe, and H. Schut). American Psychological Association, Washington, DC.

4.7

Anxiety disorders

Contents

4.7.1 Generalized anxiety disorders

Stella Bitran, David H. Barlow, and David A. Spiegel

Anxious apprehension and overconcern are common to many anxiety and mood disorders. Prior to 1980 in the American DSM diagnostic system, and 1992 in the international ICD system, individuals who experienced those symptoms in the absence of a realistic focus of concern were classified as having an 'anxiety neurosis' (DSM-II) or 'anxiety state' (ICD-9). In DSM-III, panic disorder was split off from that classification, and the residual category was renamed generalized anxiety disorder (**GAD**). A similar nomenclature was adopted in ICD-10.

Since its inception, GAD as a nosological entity has been troubled by problems of poor reliability and high comorbidity.[1] Those concerns have prompted several revisions of the DSM criteria and also have raised more basic questions regarding the validity of GAD as a disorder distinct from other anxiety and mood states. The question of what is the nature of GAD is still being debated and it remains one of the least reliably diagnosed anxiety or mood disorders.[2] This diagnostic unreliability has led to various suggestions for revisions to the diagnostic criteria and criticisms of the current definition of GAD.

Clinical features

Individuals with GAD experience persistent anxiety and worry that is out of proportion to actual events or circumstances. Typically, the anxiety and worry involve minor or everyday matters, such as work, finances, relationships, the health or safety of loved ones, and routine tasks. Often, the focus of worry shifts from one concern to another. Although people with GAD do not always consider their worries to be unrealistic or excessive, they do find them difficult to control. Consequently, the worries often interfere with concentration and performance.

Associated with the anxiety and worry, individuals with GAD have a variety of cognitive and somatic symptoms, including trembling, feeling shaky, aching in the back and shoulders, tension headaches, chest tightness, restlessness, exaggerated startle, irritability, insomnia, fatigue, dry mouth, sweating, urinary frequency, trouble swallowing, nausea, and diarrhoea. In addition, GAD may be accompanied by other conditions typically associated with stress, such as irritable bowel syndrome or atypical chest pain.

Classification

Diagnosis

(a) DSM criteria

In DSM-III, GAD was essentially a residual category for individuals with somatic symptoms of anxiety who did not meet diagnostic criteria for another, more specific, anxiety disorder. Diagnosis required the presence, for at least a month, of symptoms from three of four symptom clusters: motor tension, autonomic hyperactivity, apprehensive expectation, and vigilance and scanning. Unfortunately, clinicians had difficulty applying those criteria. In addition, its diagnosis depended on the application of the criteria for other diagnoses since GAD was not diagnosed if another anxiety disorder was present.

In DSM-III-R, apprehensive expectation was removed from the diagnostic symptom clusters, was redefined as unrealistic or excessive anxiety and worry about two or more life circumstances, and was made the essential feature of GAD. In addition, the duration criterion was changed from 1 to 6 months, and the hierarchical exclusion rule was dropped, allowing GAD to be diagnosed in addition to other disorders.

Table 4.7.1.1 DSM-IV inclusion criteria for GAD

(a) Excessive anxiety and worry, occurring more days than not for at least 6 months, about a number of events or activities
(b) The person finds it difficult to control the worry
(c) The anxiety and worry are accompanied by at least three of the following six symptoms (one in children): restlessness or feeling keyed up or on edge; being easily fatigued; difficulty concentrating or mind going blank; irritability; muscle tension; sleep disturbance
(d) The anxiety, worry or physical symptoms cause significant distress or functional impairment

(American Psychiatric Association (2000), *Diagnostic and statistical manual of mental disorders* (4th edn, text revision). APA, Washington, DC)

Despite those changes, the diagnostic reliability of GAD remained essentially unchanged.[1] Investigations revealed that the new worry criterion was problematic. Interviewers commonly disagreed as to whether two distinct spheres of worry were present, whether the worry was unrealistic or excessive, or whether the focus of the worry could be construed to be part of the symptomatology of another disorder. Moreover, studies indicated that patients with GAD did not differ substantially from control subjects in the content of their worries.[3,4] The main difference between patients and controls was that the former experienced their worrying to be uncontrollable while the latter did not.

Based on those and other findings, the GAD criteria were revised again in DSM-IV. The 'unrealistic' descriptor and the requirement for anxiety or worry to involve at least two spheres of life circumstances were deleted, and a new criterion was added that the worry must be experienced as difficult to control. In addition, the associated symptom criterion was modified to require only three of six symptoms from the previous motor tension and vigilance and scanning clusters (Table 4.7.1.1). For additional information about the evolution of the DSM criteria for GAD, see Barlow or Wincze.[5]

(b) ICD-10 criteria

Like DSM-IV, ICD-10 requires a period of 6 months of generalized anxiety and worry accompanied by certain somatic symptoms (Table 4.7.1.2). The 6 months of 'prominent' tension and worry needs to be accompanied by at least 4 of 22 associated symptoms. The ICD-10 differs from DSM-IV in that it does not require that worry be 'uncontrollable', that the symptoms of GAD occur exclusively outside the context of a mood disorder, or that they meet a 'clinical significance' criterion.

(c) Differential diagnosis

Everyone experiences anxiety and worry sometimes, and some people describe themselves as born worriers. GAD differs from these non-pathological anxiety experiences in that it is both persistent and severe enough to cause significant distress or interference. Typically also, the worries are more pervasive and difficult to control than normal worries and are associated with physical symptoms of anxiety and tension.

A number of general medical conditions can present with signs and symptoms resembling GAD (Table 4.7.1.3). In addition, substances such as caffeine, alcohol, other drugs of abuse, toxins, and some medications (Table 4.7.1.4) can cause anxiety-like

Table 4.7.1.2 ICD-10 inclusion criteria for GAD

(a) At least 6 months of prominent tension, worry, and feelings of apprehension about everyday events and problems
(b) At least four of the following 22 symptoms must be present, at least one of which must be from the autonomic arousal cluster
Autonomic arousal symptoms: palpitations or pounding heart or accelerated heart rate; sweating; trembling or shaking; dry mouth (not due to medication or dehydration)
Symptoms involving the chest and abdomen: difficulty in breathing; feeling of choking; chest pain or discomfort; nausea or abdominal distress
Symptoms involving mental state: feeling dizzy, unsteady, faint or lightheaded; derealization or depersonalization; fear of losing control, 'going crazy', or passing out; fear of dying
General symptoms: hot flushes or cold chills; numbness or tingling sensations
Symptoms of tension: muscle tension or aches and pains; restlessness and inability to relax; feeling keyed up, or on edge, or mentally tense; a sensation of a lump in the throat, or difficulty in swallowing
Other non-specific symptoms: exaggerated response to minor surprises or being startled; difficulty in concentrating, or mind 'going blank' because of worrying or anxiety; persistent irritability; difficulty getting to sleep because of worrying

(World Health Organization (2004), *International statistical classification of diseases and health related problems* (2nd edn). WHO, Geneva, Switzerland)

Table 4.7.1.3 General medical conditions that can cause symptoms resembling anxiety

Cardiac conditions: arrhythmias, coronary insufficiency, mitral valve prolapse, heart failure
Endocrine conditions: hyperthyroidism, hypoparathyroidism, hypoglycaemia
Neurological conditions: temporal lobe epilepsy, vestibular nerve disease
Respiratory conditions: asthma, hypoxia, hyperventilation, obstructive lung disease, pulmonary embolism
Other conditions: porphyria, carcinoid tumour, systemic lupus erythemacosus, pellagra

Table 4.7.1.4 Medications that can cause symptoms resembling anxiety

Psychotropics: antidepressants, neuroleptics (akathisia), sedative hypnotics (withdrawal syndrome), disulfiram
Respiratory drugs: B-adrenergic stimulants, bronchodilators
Cardiovascular drugs: antiarrhythmics, antihypertensives
Neurological disorder medications: anticonvulsants, anticholinergic agents, L-dopa
Anaesthetic drugs: pre-anaesthetics, general anaesthetics (post-anaesthetic syndrome)
Other drugs: thyroid hormone, antibiotics, non-steroidal anti-inflammatory drugs, anticancer drugs

symptoms either as a direct effect or as part of a withdrawal syndrome. These causes may be established on the basis of a medical and substance use history, physical examination, or laboratory tests.

GAD is distinguished from other psychiatric disorders, in part, by the focus of the anxiety and worry, which is not limited to a feature of another disorder. For example, the worry is not only about the possible occurrence or implications of panic attacks (as in panic disorder), or about negative evaluations by others (social phobia), gaining weight (anorexia nervosa), or having a serious illness (hypochondriasis). In obsessive–compulsive disorder, the anxiety and worry are associated with intrusive thoughts, images, or impulses that are distressing.

Generalized anxiety commonly occurs in depression, and GAD and depression also share associated symptoms such as sleep disturbance, fatigue, restlessness, and poor concentration. When the associated symptoms could fit with either disorder, the distinction is made on the basis of the presence and time course of depressed mood relative to anxiety. In DSM-IV, GAD is not diagnosed if its features occur exclusively during a mood disorder.

(d) Epidemiology

Prevalence estimates of GAD vary considerably with the diagnostic criteria used. One large-scale study found a 12-month prevalence rate of 2.07 per cent and a lifetime prevalence rate to be 4.1 per cent.[6] Socio-demographic factors associated with increased risk for GAD included being female, middle-aged, and with low income. However, being African American, Asian, or Hispanic was associated with a decreased risk.[7]

The National Comorbidity Study Replication (NCS-R), which used DSM-IV criteria and included structured interviews of over 9000 individuals in the United States, found a 12-month prevalence of GAD of 3.1 per cent and a lifetime prevalence rate at 5.7 per cent.[8,9] Lifetime prevalence rates were lowest among 18- to 29-year-olds (4.1 per cent) and those 60 or older (3.65 per cent), with the highest rates found among 45- to 59-year-olds (7.7 per cent).

(e) Comorbidity

GAD usually coexists with other anxiety and mood disorders. One large-scale study found that 68 per cent of individuals with a principal diagnosis of GAD met criteria for another Axis I disorder (Table 4.7.1.5).[10] The most frequently comorbid disorders were MDD, social phobia, or panic disorder with or without agoraphobia. Ninety-two per cent of individuals from this study with a principal diagnosis of GAD met criteria for another lifetime disorder, with 64 per cent meeting criteria for MDD. Similarly, in the major epidemiological surveys, nearly two-thirds of individuals with GAD had additional DSM Axis I diagnoses.[11] Most common among these were specific (21–59 per cent) and social (16–59 per cent) phobias, followed by panic disorder (3–27 per cent) and depression (8–39 per cent). In addition, GAD was found to be approximately twice as common among women as men. There is less information about the prevalence of personality disorders among patients with GAD.

Table 4.7.1.5 Prevalence of comorbid disorders in 279 patients with GAD

Any additional lifetime disorder	96%
Any anxiety/mood	94%
Any anxiety disorder	85%
Any mood disorder	74%
Anxiety disorders	
Panic disorder with/or without agoraphobia	47%
Social phobia	46%
Specific phobia	22%
Mood disorders	
Major depressive disorder	67%
Dysthymia	11%

(f) Is GAD a valid disorder?

The findings of only fair diagnostic reliability and high comorbidity for GAD have been interpreted as indicating poor discriminant validity of the disorder, suggesting that differentiating GAD from other anxiety and mood disorders may be artifactual. In considering those arguments, it is important to distinguish the diagnostic criteria sets specified in the DSM and ICD classification systems from the clinical syndromes they are intended to identify. Low discriminant validity for a disorder may be due to problems with the former rather than the latter. To establish the construct validity of a syndrome, one must demonstrate that it has a consistent set of features, the pattern of which separates it from other related syndromes. One approach to doing that is to compare the profiles of patients with different diagnoses across various illness dimensions.

In one such study, data from patients who took part in the DSM-IV mixed anxiety-depression field trial were examined.[12] Using factor analyses of patients' scores on 73 items from the Hamilton anxiety and depression rating scales, four clusters were identified that corresponded to the dimensions of anxiety, depression, physiological arousal, and general negative affect (containing items that loaded on both the anxiety and depression factors). Patients with a principal diagnosis of GAD had a unique profile (high on negative affect and anxiety, low on physiological arousal and depression) that distinguished them from individuals with panic disorder, major depression, anxiety or depressive disorder not otherwise specified, or no mental disorder.

A subsequent study, using an anxiety clinic sample and an expanded array of measures, yielded similar results.[13] In this case, five primary factors (corresponding to panic, agoraphobia, social anxiety, obsessions/compulsions, and general anxiety) and a higher order factor (negative affect) were identified. Again, patients with GAD had a unique factor profile.

The findings from the preceding studies were replicated and extended in an independent sample of patients.[14] As in the earlier studies, GAD was found to be distinct from other anxiety syndromes and depression, although it had the highest degree of overlap with other syndromes, especially depression. In addition, GAD was strongly associated with the non-specific dimension of negative affect, which is common to anxiety and depression. The authors suggest that GAD may represent a 'basic emotional disorder', because it consists of features that are present to varying degrees in all anxiety and mood disorders.

Finally, all three of the preceding studies (and a variety of others, e.g. Barlow *et al.*[15] support the differentiation of symptoms of autonomic arousal, which are characteristic of panic attacks, from somatic symptoms related to central nervous system tension, which form the DSM-IV-associated symptom cluster of GAD.

(g) Aetiology

Findings from genetics, neurobiology, and psychology infer a multifactorial aetiology for GAD, which has been organized into a triple vulnerabilities model.[16,17] This model suggests that anxiety disorders result from the combination of a generalized biological vulnerability, a general psychological vulnerability, and a specific psychological vulnerability.

(i) Generalized biological vulnerability

Genetic contributions. Several studies investigating genetic vulnerabilities for mental disorders have supported the notion that a shared vulnerability underlies anxiety disorders.[18] It was shown through a meta-analysis of genetic epidemiological studies that many anxiety disorders (including GAD, panic disorder, phobias, and OCD) aggregate in families and that genetics has the most influence when examining familial risk.[19] In a family study that used DSM-III criteria, GAD (but not other anxiety disorders) was five times more prevalent (19.5 per cent versus 3.5 per cent) among first-degree relatives of patients with GAD than among relatives of controls.[20] However, two twin studies using the same criteria found concordance rates for GAD were no higher among monozygotic than dizygotic twins.[21] Two subsequent studies that used DSM-III-R criteria found a shared heritability for GAD and mood disorders.[22] At present, it appears that genetic factors play a modest role in the aetiology of GAD, and one that is more closely related to vulnerability for depression than for other anxiety disorders.

Neurobiological mechanisms. A variety of neuroanatomical, neurochemical, neuroendocrine, and neurophysiological systems have been implicated in the pathogenesis of anxiety states. Much of this information has come from animal models and research on the effects of stress. Studies of neurobiological functioning in humans with GAD are limited. Some of the physical systems that may be involved in the emotion of anxiety are summarized below. Additional information may be found in reviews by Davidson,[23] and Gray and McNaughton.[24]

The noradrenergic nervous system. Noradrenergic pathway (the **locus coeruleus-noradrenaline-sympathetic nervous system**) have long been associated with fear and arousal and play an important role in the body's response to threat. However, their role in persistent anxiety states is not clear. Resting catecholamine levels in patients with GAD appear to be normal. On the other hand, GAD patients exhibit subnormal responses to both stimulation[25] and blockade[26] of α_2-adrenergic receptors and a reduced density of α_2-receptors in platelets.[27] Those findings could reflect downregulation of the α_2-receptors due to initially high levels of noradrenaline (norepinephrine).

Consistent with those neurochemical findings, somatic measures of autonomic nervous system function (e.g. skin conductance, respiratory rate, heart-rate variability, blood pressure) in patients with GAD tend to show normal resting values with blunted and sometimes prolonged responses to stressful stimuli.[27] Psychophysiological studies have found that worry is associated with restricted sympathetic arousal and low vagal tone.[27] In contrast, it has been shown that compared to controls, individuals with GAD show greater muscle tension at baseline in response to psychological challenge.[28] In addition, structural analyses suggest that GAD, unlike other anxiety disorders, is not associated with autonomic hyperarousal when levels of negative affect are held constant.[14] These findings indicate diminished autonomic nervous system responsiveness in individuals with GAD.

The hypothalamic-pituitary-adrenal axis. The **hypothalamic-pituitary-adrenal axis** and its end-product, cortisol, are also involved in reactions to stress. Activity in the hypothalamic-pituitary-adrenal axis is subject to a variety of influences. Primary control is by means of hypothalamic secretion of corticotrophin-releasing factor, which stimulates pituitary secretion of ACTH, which in turn stimulates adrenal secretion of cortisol. Circulating cortisol, and analogues such as dexamethasone, exert inhibitory feedback at the level of the pituitary gland and apparently also by means of receptors on the hippocampus.

In rats, chronic exposure to stress or exogenous steroids results in a reduction of corticosteroid receptors in the hippocampus and a consequent decrease in feedback inhibition by cortisol.[29] These animals exhibit reduced dexamethasone suppression of cortisol secretion and greater or more prolonged adrenocortical responses to stress. Reduced dexamethasone suppression also has been observed in approximately one-third of patients with DSM-III-diagnosed GAD.[30] This reduction in the normal regulatory control of cortisol secretion may be one mechanism through which chronic or repeated stress can lead to persistent anxiety.

The amygdala and the bed nucleus of the stria terminalis. LeDoux[31] and others have demonstrated the central role played by the **amygdala** in the mediation of fear reactions. The amygdala is thought to be responsible for the detection of potential threats to the organism and the mobilization of a range of defensive responses (Fig. 4.7.1.1). Through connections with the hypothalamus, it can activate the sympathetic nervous system and hypothalamic-pituitary-adrenal axis. Through efferent fibres to the central grey area of the midbrain, it can mediate behavioural defence responses such as the fight-or-flight response and behavioural 'freezing'. Through connections to the nucleus reticularis pontis caudalis, it can enhance the defensive startle reflex.

A structural magnetic resonance imaging (MRI) study of children and adolescents found that those individuals with GAD had increased total and right amygdala volume compared to non-anxious controls.[32] In addition, abnormalities in fear circuitry, especially hyperactivation in the right amygdala, have been found in adolescents with GAD.[33]

The extent to which these pathways are involved in the neurobiology of anxiety (as opposed to fear) is unclear. However, a structure

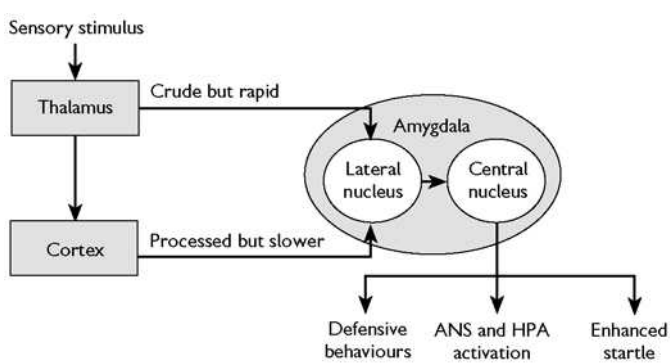


Fig 4.7.1.1 Fear pathways (based on descriptions by LeDoux[31] and Davis[34]). ANS, autonomic nervous system; HPA, hypothalamic-pituitary-adrenal axis.

closely related to the amygdala, the bed nucleus of the stria terminalis, may be involved in this emotion. The bed nucleus resembles the amygdala in its neurotransmitter content, cell morphology, and hypothalamic and brainstem connections and, like the amygdala, it exerts a modulating effect on the startle reflex.[34] Studies of this latter effect implicate it in the experience of anxiety.

Administration of corticotrophin-releasing factor into the cerebral ventricles of rats produces a state of generalized arousal resembling anxiety. Under those conditions, the startle reflex also is enhanced. Exposing rats to bright light for 5 to 20 min has similar effects. These effects are not blocked by damage to amygdala but are by lesions to the bed nucleus of the stria terminalis and by treatment with benzodiazepines or buspirone. Conversely, infusion of corticotrophin-releasing factor directly into the bed nucleus of the stria terminalis, but not the amygdala, produces a rapid increase in startle. Based on these observations, Davis[34] has suggested that the stria terminalis may play a role in anxiety analogous to that of the amygdala in fear reactions and, further, that prolonged or repeated stimulation of the stria terminalis by corticotrophin-releasing factor during periods of stress might lead to sustained activation and thus to persistent anxiety. A recent study has confirmed the differential association of these structures with fear and anxiety, respectively.[35]

The septohippocampal system (behavioural inhibition system). The bed nucleus of the stria terminalis is part of the larger septohippocampal system.[36] In 1982, based on data from several lines of research, Gray hypothesized that the septohippocampal system, together with the Papez circuit (a neural loop connecting the subicular area in the hippocampal formation to the mammillary bodies, anterior thalamus, cingulate cortex, and back to the subiculum), is responsible for mediating the emotion of anxiety as well as the major effects of anxiolytic drugs.[36] Gray called this network the **behavioural inhibition system**, because he believed that, when activated, it interrupts ongoing behaviour and redirects the organism's attention to signs of possible danger.

According to Gray's model,[24,36] the behavioural inhibition system receives information about the environment from the sensory cortex via the temporal lobe and hippocampal formation. The system checks the information for consistency with predictions, which are updated continuously by the Papez circuit based on preceding information and stored patterns, as well as for consistency with the immediate goals of the organism. When a mismatch is found, or if a predicted event is aversive, the outputs of the behavioural inhibition system are activated, resulting in a constellation of emotional and behavioural effects consistent with anxiety (Fig. 4.7.1.2).

The activation of the behavioural inhibition system appears to be moderated by ascending noradrenergic and serotonergic projections to the septohippocampal complex, providing a possible mechanism for the anxiolytic actions of some drugs. The amygdala also provides inputs to the behavioural inhibition system and may relay its outputs to the hypothalamus and autonomic nervous system, thereby mediating anxious arousal. Sustained activation of the behavioural inhibition system might therefore account for many of the features of GAD.

The benzodiazepine-γ-aminobutyric acid system. The powerful anxiolytic and sedative effects of benzodiazepines are believed to be mediated by benzodiazepine recognition sites located on γ-aminobutyric acid (**GABA**) type A receptor complexes in the central nervous system. When bound to those complexes,

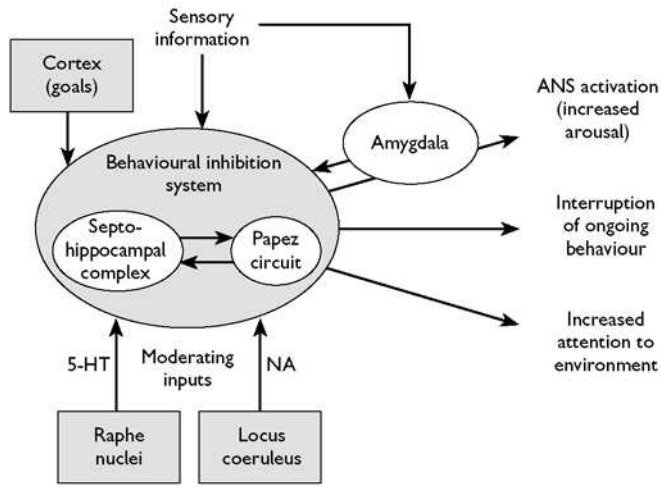


Fig 4.7.1.2 Behavioural inhibition system (based on descriptions by Gray and McNaughton[24]). ANS, autonomic nervous system; 5-HT, serotonin; NA, noradrenaline.

benzodiazepines allosterically modulate the GABA receptors to enhance the normal inhibitory effects of GABA on neurotransmission. Activation of **central benzodiazepine-GABA receptor complexes** also suppresses hypothalamic-pituitary-adrenal axis activity and, consequently, cortisol levels.

In addition to these central receptor complexes, benzodiazepine recognition sites of a different type are present widely in cells outside the central nervous system. These so-called **peripheral benzodiazepine receptors** are believed to be instrumental in controlling the synthesis of regulatory steroids. Their role in the anxiolytic actions of benzodiazepines is unknown; although they bind some drugs (e.g. diazepam), they have low affinity for others (e.g. clonazepam). Interestingly, peripheral benzodiazepine receptors are decreased in blood cells of individuals with untreated GAD but return to normal levels after successful treatment with benzodiazepines. Their numbers also vary in response to stress, being elevated following acute stressors and reduced during chronic stress.

A possible explanation for those changes has been suggested by Rocca *et al.*[37] The investigators note that peripheral benzodiazepine receptors in brain glial cells control the production of neurosteroids that act as modulators of $GABA_A$ receptor sensitivity. Their effect on GABA functioning appears to be opposite to that of clinically effective benzodiazepines, that is, they 'decrease' rather than increase the inhibitory effects of GABA. It is hypothesized that an endogenous ligand of these glial cell receptors (possibly diazepam-binding inhibitor) is released during stress, initiating the cycle of events depicted in Fig. 4.7.1.3.

The immediate effect of these events would be to enhance the stress-induced release of cortisol. However, prolonged cortisol excess is hypothesized to downregulate peripheral benzodiazepine receptors, resulting in the reduced receptor densities found in GAD. Administration of a clinically effective benzodiazepine drug would interrupt the proposed pathway at the point of the central GABA receptor, lowering cortisol levels and restoring synthesis of peripheral benzodiazepine receptors.

Other neurotransmitter systems. Individuals with GAD have been reported to have reduced serotonin levels in the cerebral spinal fluid[35] and decreased platelet binding of paroxetine, a selective serotonin reuptake inhibitor.[38] In addition, drugs that affect

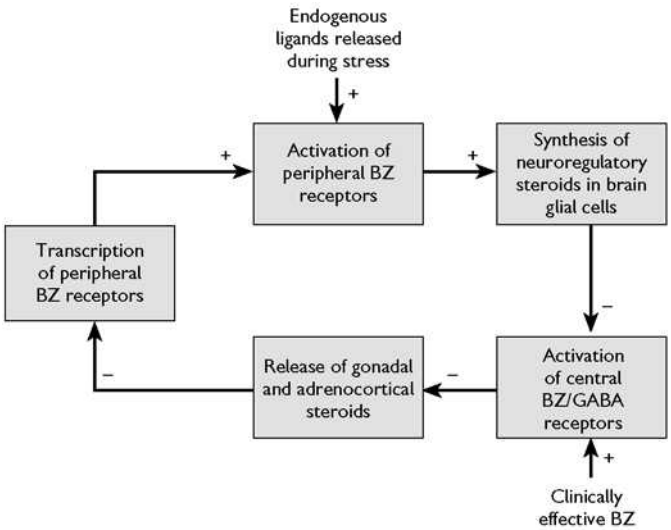


Fig 4.7.1.3 Possible involvement of peripheral benzodiazepine receptors in acute and chronic stress reactions (based on descriptions by Rocca *et al.*[37]). BZ, benzodiazepine.

serotonergic transmission (e.g. buspirone and venlafaxine) are effective in the treatment of GAD. These findings suggest that serotonin regulation may be abnormal in GAD.

Cholecystokinin neuropeptides (CCK-4 and CCK-8S) have been implicated in the genesis of arousal and fear responses.[40] It is unclear how those effects are mediated; however, cholecystokinin interacts with several neurotransmitters and systems believed to be involved in anxiety responses, including the noradrenergic nervous system, the hypothalamic-pituitary-adrenal axis, the benzodiazepine-GABA system, and serotonin.

(ii) Generalized psychological vulnerability

A diminished sense of control. Early experiences of uncontrollability may serve as a psychological vulnerability for emotional disorders.[16] Individuals in clinical populations, including people diagnosed with anxiety disorders, sexual dysfunctions, and depression, often perceive themselves as having little control over their experiences.[41,42] Patients with GAD are more likely than controls to perceive a lack of control over threatening events and to regard ambiguous information as threatening.[43] This perceived lack of control may result from a variety of events, including trauma and insecure attachment to primary caregivers.[44] In patients with GAD, worry may be an ineffective attempt to assert control on uncertain future events. Intolerance of uncertainty, a construct related to perceived lack of control, has emerged as an important variable in the study of anxiety disorders.[45] Defined as the inability to accept that future negative events may occur, intolerance of uncertainty has been associated with symptoms of several anxiety disorders, but is greater in individuals with GAD than in patients from a mixed anxiety disorder sample.[41]

Parenting. There is an extensive literature on the influences of early environmental factors on the development of anxiety and other negative emotions in children (for an integrative review, see Chorpita and Barlow[46]). Attachment theory holds that parents or other consistent caregivers serve an important function in a child's development by providing a protective and secure base from which the child can operate. Disruption of this base is hypothesized to lead initially to anxious apprehension and dependency and, if the disruption is severe, subsequently to withdrawal and depression.

An important aspect of a healthy parent–child relationship is its ability to foster in the child a sense of control over events. According to Chorpita and Barlow,[46] an individual who lacks sufficient early experiences of control may develop a general perception of personal inefficacy, which may predispose him or her to chronic negative emotional states such as GAD. Two aspects of parenting appear to be important in providing a child with opportunities to experience control: responsiveness to the child's efforts at engagement and encouragement of the child to explore and manipulate the environment. A parenting style characterized by excessive control of the child's environment (overprotection) coupled with a lack of warmth and responsiveness toward the child would deprive the child of such opportunities and thus, theoretically, could contribute to the development of anxiety.

Consistent with this theory, mothers of anxious preschool children were found to be more critical and intrusive and less responsive to their children than mothers of non-anxious children.[47] In addition, adults who rated their parenting as more protective and less caring had higher trait anxiety scores than other individuals surveyed. A similar pattern was found to distinguish patients meeting DSM-III-R criteria for GAD or panic disorder from controls.[48] One hypothesis is that the relationship of these early parenting experiences to the subsequent development of anxiety (or depression) is mediated by the early formation of cognitive vulnerability best described as a sense of uncontrollability regarding future events in one's life (Fig. 4.7.1.4).[46]

Specific psychological vulnerability. Data from twin studies suggest that environmental influences contribute more to the variance in aetiology of GAD than heredity.[19] In addition, the triple vulnerability model of the aetiology of anxiety disorders suggests that psychological and biological vulnerabilities interact with

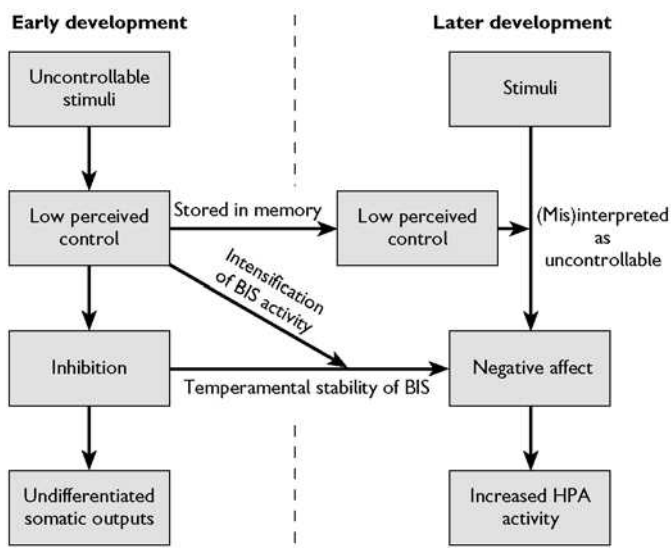


Fig 4.7.1.4 Model of the development of vulnerability for anxiety and depression. BIS, behavioural inhibition system[24]; HPA, hypothalamic-pituitary-adrenal axis (Copyright © 1998 by the American Psychological Association. Reproduced with permission. Chorpita, B.F. and Barlow, D.H. The Development of anxeity: the role of control in the early environment, *Psychological Bulletin*, **124**, 3–2. The use of APA information does not imply endorsement by APA).

environmental factors (i.e. stressful life events), to produce anxiety symptoms that reach a clinical level.[16,17] Several environmental factors have been implicated.

Stressful life events. Several studies have found an association between stressful or traumatic life events and GAD. The natures of the stressors that precede the onset of GAD remain unclear. However, in one study, 52 per cent of individuals diagnosed with GAD reported experiencing at least one past traumatic event (i.e. an event that would meet for Criterion A of the DSM-IV definition of PTSD), compared to only 21 per cent of non-anxious controls. It was unclear in this study if the events occurred prior to the onset of the GAD.[49] In addition, a variety of stressors have been associated with increased risk of GAD, including early parental death,[50] rape or combat,[51] and chronically dysfunctional marital and family relationships.[52]

Course and prognosis. Although there is evidence to the contrary, GAD has often been considered a chronic and disabling condition. The Harvard/Brown Anxiety Research Program study provides information on the course and impact of GAD among patients treated naturalistically over a 3-year period.[53] The mean age at onset of GAD was 21 years (range 2–61 years), and the average duration of illness at initial evaluation was 20 years. Excluding patients with comorbid panic disorder, one-third of subjects had never married and another 17 per cent were separated, widowed, or divorced. Unemployment was higher than average, and 37 per cent of subjects had received public financial assistance. Despite the fact that more than 80 per cent of patients received treatment during the study period, remission from GAD was uncommon (15 per cent at 1 year, 27 per cent by 3 years). Among patients with comorbid psychiatric disorders, the proportions achieving remission from GAD and coexisting anxiety disorders were only 8 per cent and 17 per cent at 1 and 3 years, respectively.

However, there is evidence that the perception of GAD as a chronic, unremitting condition may not be completely correct. A longitudinal study of individuals with GAD found that many (46 per cent of women and 56 per cent of men) experienced episodes of full remission and that the periods with no symptoms last longer in women.[54]

Treatment

(a) Pharmacotherapy

Several pharmacological agents have been shown to be effective for the treatment of GAD (for a review, see Davidson[23]). Chief among these are the benzodiazepines, azapirones, and antidepressants. Evidence shows that several types of medications may be effective for at least short-term relief of anxiety, however, many are not effective in the long term unless taken indefinitely.[55]

(i) Benzodiazepines

Benzodiazepines have for decades been prescribed for short-term relief of anxiety.[55] Although, evidence demonstrates that these drugs can be effective in relieving anxiety for a short period, there is little or no evidence that they work over a long period. However, compared with other agents used to treat anxiety disorders, they are safe, fast-acting, and have relatively few side-effects. All currently available benzodiazepines probably are efficacious for GAD. Approximately two-thirds of patients experience moderate to marked improvement, with effects being evident within the first 1 or 2 weeks of treatment.[56]

Benzodiazepines appear to be more effective for the somatic symptoms of GAD than for psychic symptoms such as apprehensive worry and irritability, possibly because of their sedative and myorelaxant properties. In some studies, irritability actually has increased during treatment with benzodiazepines. Consequently, these drugs may be better for patients whose complaints are more somatic than psychic, whereas other agents may be better when the reverse is true.

(ii) Azapirones

In recent years, azapirone drugs have become a popular alternative to benzodiazepines for the treatment of GAD. These drugs lack the sedative and muscle relaxant properties of the benzodiazepines as well as their ability to potentiate the effects of alcohol. However, improvement is somewhat slower (2–4 weeks) than with benzodiazepines. The most widely used of these agents is buspirone, whose efficacy and safety have been demonstrated in several well-controlled trials (see Rickels[57] for a review) and became, in 1996, the only non-benzodiazepine approved by the U.S. Food and Drug Administration for the treatment of GAD.[58] It is as effective as benzodiazepines for general anxiety and may reduce some of the associated features of GAD as well, including depression and agitation. Buspirone also does not seem to be associated with as much dependence and withdrawal as found with the benzodiazepines.[55] Ipsapirone, an azapirone with somewhat greater affinity and selectivity for the 5-hydroxytryptamine-1A receptor and fewer side-effects than buspirone also appears to be effective.[59] Other drugs in this class include gepirone, tandospirone, and flesinoxan.

(iii) Antidepressants

Both imipramine and trazodone have been superior to placebo for the treatment of GAD in controlled trials. In one trial, the two drugs were comparable to each other and to diazepam in reducing anxiety after the first 2 weeks of treatment.[56] In another study, imipramine was as effective as chlordiazepoxide overall and produced greater reductions in associated depression. In addition, nefazodone, an agent related to trazodone but less sedating, was effective for GAD in a small open trial.[60]

Many studies have examined the use of newer antidepressants, including selective reuptake inhibitors (SSRIs) and serotonin and norepinephrine reuptake inhibitors (SRNIs), in the treatment of GAD and have had favourable results. One placebo-controlled study of the SSRI paroxetine found that 62 per cent and 68 per cent of patients receiving 20 and 40 mg of paroxetine had significant reduction in symptoms, compared to only 46 per cent in the placebo group.[61] In addition, due to results from several large, placebo-controlled trials, venlafaxine was the first antidepressant approved by the Federal Drug Administration for the treatment of GAD.[62,63]

(b) Psychosocial treatment

Early psychological treatments for GAD consisted mostly of non-specific interventions such as supportive psychotherapy, relaxation training, and biofeedback. In general, those treatments were not very effective. However, cognitive behavioural treatment (CBT) has been found in several randomized controlled trials to be associated with clinically significant improvement. Cognitive behavioural treatment consists of psychoeducation about the nature of anxiety,

symptom monitoring, relaxation training, exposure (imaginal and *in vivo*), and cognitive restructuring. Psychoeducation is a common CBT component in which the nature of anxiety is discussed in order to normalize the patient's experience. It also serves to teach patients to differentiate between adaptive and maladaptive anxiety. Symptom monitoring allows records of mood and anxiety to be recorded and allows patients to learn to observe their symptoms. Progressive muscle relaxation (PMR) is a common technique used to teach patients relaxation strategies to use in anxiety-provoking situations. Exposures are tailored to the nature of GAD and are either *in vivo* or imaginal and are often combined with rehearsal of coping skills (i.e. PMR) until anxiety subsides. Cognitive restructuring targets distorted cognitions and information-processing biases associated with GAD. Strategies to challenge catastrophizing or overestimating the probability of a negative event are often employed during this component of treatment. CBT is typically administered in a dozen or so sessions, and can be conducted in group or individual formats. In controlled trials, cognitive behavioural treatments have been more effective than no treatment or a psychological or drug placebo treatment and at least as effective as benzodiazepines (for a review, see Barlow *et al.*).[64] Attrition is low (10–15 per cent), and reductions in anxiety average about 50 per cent, with gains being maintained at follow-up. Currently, the most successful treatments combine relaxation training with cognitive interventions focused on making the worry process more controllable.

Consistent with the findings of individual studies, meta-analyses for GAD have found that CBT is more efficacious than control conditions, resulting in medium to large effect sizes when compared to pill or psychological placebo.[65] However, in studies directly comparing CBT with pharmacotherapy (most often benzodiazepines), there was no difference in effect sizes between the two treatments. Currently, neither mode of treatment has been shown to be consistently superior to the other.

Recent developments in psychosocial treatments for GAD have integrated acceptance and mindfulness approaches into traditional CBT.[66] These treatments are based on the notion that individuals' attempts to control their internal experience often backfire, resulting in increased anxiety. Components to these treatments may include experiential exercises, mindfulness training, identification of overriding values, and encouragement towards taking action in ways that are consistent with these values.[66] Further exploration with controlled trials needs to be conducted to examine the efficacy of these newer approaches with cognitive behavioural treatments.

(c) Combined pharmacotherapy and psychotherapy

Although common in clinical practice, little is known about the effects of combining pharmacotherapy and psychotherapy for GAD. In the only published study to date, Power *et al.*[67] compared cognitive behavioural therapy, diazepam, a pill placebo, cognitive behavioural therapy plus diazepam, and cognitive behavioural therapy plus a pill placebo in a sample of DSM-III-diagnosed GAD patients. At post-treatment and follow-up, patients in all three cognitive behavioural therapy conditions were more improved than those who received diazepam or placebo alone. Although the cognitive behavioural therapy groups did not differ significantly from each other on any measure, the cognitive behavioural therapy plus diazepam group improved earliest and had the largest percentage of patients achieving a criterion of clinically significant change on all measures. Unfortunately, the use of DSM-III criteria in this study makes its relevance to GAD as it is currently defined uncertain. In addition, the cognitive behavioural therapy used was briefer (seven sessions) and less specific than the currently recommended forms.

(d) Effect of comorbidity on treatment outcome and vice versa

Many treatment trials investigating GAD have excluded patients with comorbid Axis I disorders. A review of 48 GAD studies published between 1980 and 1991 found that only eight reported including patients with other psychiatric disorders.[68] When comorbid disorders have been permitted, their effect on treatment outcome generally has not been evaluated. In the Harvard-Brown Anxiety Research Program study, the presence of a comorbid psychiatric disorder reduced response rates at 1 and 3 years by nearly 50 per cent. In addition, it has been found that concurrent personality disorders impair outcome of treatment for GAD. In one study improvement was comparable among treatment completers with or without Axis II disorders, but attrition was greater in the former (44 per cent) than the latter (19 per cent) group.[69]

The high rates of comorbidity associated with GAD have important implications for treatment. Higher rates of comorbidity are associated with lower rates of remission and greater likelihood of relapse over 12-year follow-up.[70] On the other hand, successful treatment of GAD in patients with comorbid disorders often reduces the severity of the other disorders as well. Borkovec *et al.*[68] examined the effect of various psychosocial treatments for GAD on coexisting anxiety and mood disorders (except panic disorder or severe depression, which were excluded from the study). Across treatments, patients whose GAD improved exhibited reductions as well in other anxiety disorders (mostly social and simple phobia) and dysthymia. Only 4 of 13 successfully treated patients who had additional disorders at pre-treatment continued to have them at post-treatment.

Clinical management

Many anxious patients do not meet diagnostic criteria for GAD. These patients often respond to conservative measures. If the symptoms are minor or are related to a situational stressor, brief psychotherapy and support is the treatment of first choice. In one study, patients who initially reported physical or minor emotional complaints responded better to counselling than to diazepam, even when counselling was limited to only 3 h.[72] Often, an explanation of the relationship of physical symptoms to stress is reassuring to patients and can interrupt a spiral of symptoms leading to anxiety and worry about health, leading to increased symptoms, and so on. Simple behavioural interventions such as relaxation training for patients with prominent muscle tension or breathing exercises for those with dyspnoea or hyperventilation may be helpful as well.

For patients with marked adrenergic symptoms or insomnia, the temporary (a few days to a few weeks) use of a benzodiazepine may be helpful as an adjunct to psychotherapy. In some cases, a hypnotic drug alone is sufficient. In general, as-needed use of benzodiazepines should be discouraged, because it is more likely than scheduled use to foster reliance on drugs as the principal means of coping with anxiety. For the same reason, the drug dose should be kept as low as possible and should be tapered as therapy proceeds.

For patients meeting diagnostic criteria for GAD, treatment with an empirically validated form of psychosocial therapy for GAD is strongly recommended. Such treatments are available in manualized form with accompanying patient workbooks.[55]

When medication treatment is preferred, a trial of buspirone is a good initial choice. Exceptions are patients with comorbid panic disorder or marked adrenergic symptoms, for which benzodiazepines may be better if they are not contraindicated (see below). The typical starting dose of buspirone is 15 mg/day in divided doses (5 mg thrice daily or 7.5 mg twice a day), which is increased by 5 mg/day every few days to a target dose of 30 mg/day. If the response is insufficient after 2 to 4 weeks at that amount, or if the patient is experiencing significant depressive symptoms, the dose may be advanced gradually to a maximum of 60 mg/day. Improvement may continue for up to 3 months.

It is important to inform patients of the typical side-effects and response time of buspirone. Patients who have taken benzodiazepines previously may expect prompt relief and sedative side-effects from the medication and may become discouraged when these are absent. When switching from a benzodiazepine to buspirone, it may be helpful to continue the benzodiazepine during the first month of buspirone therapy before initiating a gradual taper. Remember that buspirone will not prevent benzodiazepine withdrawal symptoms.

Failing a course of buspirone, or in patients with comorbid major depression, trials of antidepressant medications (e.g. imipramine, venlafaxine, trazodone) would be a reasonable second choice. Venlafaxine may be effective at doses as low as 75 mg/day (the usual starting dose). Because of its short metabolic half-life, the extended release form is preferred, which allows once per day dosing and thus may improve compliance. The dose typically is advanced by 75 mg/day every 2 weeks to a maximum of 225 mg/day. Dosing for other antidepressants is the same as for the treatment of depression.

Because of the risk of dependence and (uncommonly) abuse, long-term use of benzodiazepines generally is reserved for patients who do not respond sufficiently to other options. Relative contraindications include a need to be alert (e.g. drivers and machinery operators), a personal or family history of alcoholism or drug abuse, and prominent aggressiveness or irritability. Generally, longer half-life benzodiazepines (e.g. diazepam, clonazepam), which can be taken once or twice daily, are preferred. A typical starting dose is 5 to 10 mg/day of diazepam or equivalent, which is advanced every few days to a maximum of 40 mg/day.

In instances where pharmacotherapy is the primary treatment, responders should be continued on medication for at least 6 months before a gradual drug taper is attempted. Even so, a substantial proportion will relapse after drug discontinuance and will require further treatment. Discontinuing pharmacotherapy in the context of effective psychosocial treatment may improve success rates.

Conclusions

Based on current knowledge, GAD seems to be the exaggerated expression of the human potential to apprehensively anticipate and prepare for future misfortune. As such, it may represent a 'basic' disorder, a better understanding of which may shed light on other anxiety and mood disorders. Although the definition of GAD has been revised considerably since being given status of a full disorder in DSM-III-R, much more information needs to be gathered regarding psychosocial treatments for GAD. In addition, due to its poor diagnostic reliability, many revisions to the diagnostic criteria of GAD have been suggested. Despite the progress in elucidating the nature of GAD, further research needs to be conducted to continue to enhance our understanding and development of effective and generalizable treatments.

Further information

Barlow, D.H. (1987). The classification of anxiety disorders. In *Diagnoses and classification in psychiatry: a critical appraisal of DSM-III* (ed. G.L. Tischler), pp. 223–42. Cambridge University Press, New York.

Brown, T.A., Barlow, D.H., and Liebowitz, M.R. (1994). The empirical basis of generalized anxiety disorder. *The American Journal of Psychiatry*, **151**, 1272–80.

Kessler, R.C., McGonagle, K.A., Zhao, S., *et al.* (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Archives of General Psychiatry*, **51**, 8–19.

Barlow, D.H., Chorpita, B.F., and Turovsky, J. (1996). Fear, panic, anxiety, and disorders of emotion. In *Nebraska symposium on motivation. Perspectives on anxiety, panic, and fear*, Vol. 43 (ed. D.A. Hope), pp. 251–328. University of Nebraska Press, Lincoln, NE.

Barlow, D.H. (2000). Unraveling the mysteries of anxiety and its disorders from the perspective of emotion theory. *The American Psychologist*, **55**, 1247–63.

Suarez, L., Bennett, S.M., Goldstein, C.M., *et al.* (2008). Understanding anxiety disorders from a "triple vulnerability" framework. In *Handbook of anxiety and the anxiety disorders* (eds. M.M. Antony and M.B. Stein). Oxford University Press, New York.

LeDoux, J. (1996). *The emotional brain: the mysterious underpinnings of emotional life*. Simon and Schuster, New York.

Craske, M.G. and Barlow, D.H. (2006). *Mastery of your anxiety and worry* (2nd edn). Oxford University Press, New York.

Barlow, D.H., Allen, L.B., and Basden, S.L. (2007). Psychosocial treatments for panic disorders, phobias, and generalized anxiety disorder. In *A guide to treatments that work* (3rd edn) (ed. P.E. Nathan and J.M. Gorman), Oxford University Press, New York.

References

1. DiNardo, P.A., Moras, K., Barlow, D.H., *et al.* (1993). Reliability of DSM-III-R anxiety disorder categories: using the Anxiety Disorders Interview Schedule-Revised (ADIS-R). *Archives of General Psychiatry*, **50**, 251–6.
2. Brown, T.A., DiNardo, P.A., Lehman C.L., *et al.* (2001). Reliability of DSM-IV anxiety and mood disorders: implications for classification of emotional disorders. *Journal of Abnormal Psychology*, **110**, 49–58.
3. Abel, J.W. and Borkovec, T.D. (1995). Generalizability of DSM-III-R GAD to proposed DSM-IV criteria and cross validation of proposed changes. *Journal of Anxiety Disorders*, **9**, 303–15.
4. Craske, M.G., Rapee, R.M., Jackel, L., *et al.* (1989). Qualitative dimensions of worry in DSM-III-R generalized anxiety disorder subjects and nonanxious controls. *Behavior Research and Therapy*, **27**, 189–98.
5. Barlow, D.H. and Wincze, J. (1998). DSM-IV and beyond: what is generalized anxiety disorder? *Acta Psychiatrica Scandinavica*, **98** (Suppl. 393), 23–9.
6. Grant, B.F., Stinson, F.S., Dawson, D.A., *et al.* (2004). Prevalence and co-occurrence of substance abuse disorders and independent mood and anxiety disorders. *Archives of General Psychiatry*, **61**, 807–16.
7. Grant, B.F., Hasin, D.S., Stinson, F.S., *et al.* (2005). Prevalence, correlates, co-morbidity, and comparative disability of DSM-IV

generalized anxiety disorder in the US: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychological Medicine*, **35**, 1747–59.
8. Kessler, R.C., Berglund, P., Demler, O., *et al.* (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**, 593–602.
9. Kessler, R.C., Chiu, W.T., Demler, O., *et al.* (2005). Prevalence, severity, comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**, 617–27.
10. Brown, T.A., Campbell, L.A., Lehman, C.L., *et al.* (2001). Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. *Journal of Abnormal Psychology*, **110**, 585–99.
11. Wittchen, H.-U., Zhao, S., Kessler, R.C., *et al.* (1994). DSM-III-R generalized anxiety disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, **51**, 355–64.
12. Zinbarg, R.E., Barlow, D.H., Liebowitz, M., *et al.* (1994). The DSM-IV field trial for mixed anxiety-depression. *American Journal of Psychiatry*, **151**, 1153–62.
13. Zinbarg, R. and Barlow, D.H. (1996). Structure of anxiety and the anxiety disorders: a hierarchial model. *Journal of Abnormal Psychology*, **105**, 181–93.
14. Brown, T.A., Chorpita, B.F., and Barlow, D.H. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology*, **107**, 179–92.
15. Barlow, D.H., Chorpita, B.F., and Turovsky, J. (1996). Fear, panic, anxiety, and disorders of emotion. In *Nebraska Symposium on Motivation.* Vol. 43, *Perspectives on anxiety, panic, and fear* (ed. D.A. Hope), pp. 251–328. University of Nebraska Press, Lincoln, NE.
16. Barlow, D.H., (2002). *Anxiety and its disorders* (2nd edn). Guilford Press, New York.
17. Suarez, L., Bennett, S.M., Goldstein, C.M., *et al.* (2008). Understanding anxiety disorders from a "triple vulnerability" framework. In *Handbook of anxiety and the anxiety disorders* (eds. M.M. Antony and M.B. Stein). Oxford University Press, New York.
18. Hettema, J.M., Prescott, C.A., Myers, J.M., *et al.* (2005). The structure of genetic and environmental risk factors for anxiety disorders in men and women. *Archives of General Psychiatry*, **62**, 182–8.
19. Hettema, J.M., Neale, M.C., and Kendler, K.S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Archives of General Psychiatry*, **158**, 1568–78.
20. Noyes, R., Clarkson, C., Crowe, R.R., *et al.* (1987). A family study of generalized anxiety. *The American Journal of Psychiatry*, **144**, 1019–24.
21. Andrews, G., Stewart, S., Allen, R., *et al.* (1990). The genetics of six neurotic disorders: a twin study. *Journal of Affective Disorders*, **19**, 23–9.
22. Kendler, K.S., Neale, M.C., Kessler, R.C., *et al.* (1992). Major depression and generalized anxiety disorder: same genes, (partly) different environments? *Archives of General Psychiatry*, **49**, 716–22.
23. Davidson R.T. (2001). Pharmacotherapy of generalized anxiety disorder. *The Journal of Clinical Psychiatry*, **62**(Suppl. 11), 46–50.
24. Gray, J.A. and McNaughton, N. (1996). The neuropsychology of anxiety: a reprise. In *Nebraska symposium on motivation. Perspectives on anxiety, panic, and fear*, Vol. 43 (ed. D.A. Hope), pp. 61–134. University of Nebraska Press, Lincoln, NE.
25. Charney, D.S., Woods, S.W., Heninger, G.R., *et al.* (1989). Noradrenergic function in generalized anxiety disorder: effects of yohimbine in healthy subjects and patients with generalized anxiety disorder. *Psychiatry Research*, **27**, 173–82.
26. Abelson, J.L., Glitz, D., Cameron, O.G., *et al.* (1991). Blunted growth hormone response to clonidine in patients with generalized anxiety disorder. *Archives of General Psychiatry*, **48**, 157–62.
27. Cameron, O.G., Smith, C.B., Lee, M.A., *et al.* (1990). Adrenergic status in anxiety disorders: platelet alpha two-adrenergic receptor binding, blood pressure, pulse, and plasma catecholamines in panic and generalized anxiety disorder patients and in normal subjects. *Biological Psychiatry*, **28**, 3–20.
28. Hazlett, R.L., McLeod, D.R., and Hoehn-Saric, R. (1994). Muscle tension in generalized anxiety disorder: elevated muscle tonus or agitated movement? *Psychophysiology*, **31**, 189–95.
29. Sapolsky, R.M. (2007). Stress, stress-related disease, and emotional regulation. In *Handbook of emotional regulation* (ed. J.J. Gross). Guilford Press, New York.
30. Tiller, J.W.G., Biddle, N., Maguire, K.P., *et al.* (1988). The dexamethasone suppression test and plasma dexamethasone in generalized anxiety disorder. *Biological Psychiatry*, **23**, 261–70.
31. LeDoux, J. (1998). Fear and the brain: where have we been, and where are we going? *Biological Psychiatry*, **44**, 1229–38.
32. DeBellis, M.D., Casey, B.J., Dahl, R.E., *et al.* (2000). A pilot study of amygdala volumes in pediatric generalized anxiety disorder. *Biological Psychiatry*, **48**, 51–7.
33. McClure, E.B., Monk, C.S., Nelson, E.E., *et al.* (2007). Abnormal attentional modulation of fear circuit function in pediatric generalized anxiety disorder. *Archives of General Psychiatry*, **64**, 97–106.
34. Davis, M. (1998). Are different parts of the extended amygdala involved in fear versus anxiety? *Biological Psychiatry*, **44**, 1239–47.
35. Waddell, J., Morris, R.W., and Bouton, M.E. (2006). Effect of bed nucleus of the stria terminalis lesions on conditioned anxiety: aversive conditioning with long-duration conditional stimuli and reinstatement of extinguished fear. *Behavioral Neuroscience*, **120**, 324–36.
36. Gray, J.A. (1982). *The neuropsychology of anxiety*. Oxford University Press, New York.
37. Rocca, P., Beoni, A.M., Eva, C., *et al.* (1998). Peripheral benzodiazepine receptor messenger RNA is decreased in lymphocytes of generalized anxiety disorder patients. *Biological Psychiatry*, **43**, 767–73.
38. Brewerton, T.D., Lydiard, R.B., Johnson, M.R., *et al.* (1995). CSF serotonin: diagnostic and seasonal differences. In *New research abstracts of the 148th meeting of the American Psychiatric Association*, Abstract NR385:151. American Psychiatric Press, Washington, DC.
39. Iny, L.J., Pecknold, J., Suranyi-Cadotte, B.E., *et al.* (1994). Studies of neurochemical link between depression, anxiety, and stress from [^{3}H]imipramine and [^{3}H]paroxetine binding on human platelets. *Biological Psychiatry*, **36**, 281–91.
40. Bradwejn, J., Koszycki, D., Couetoux du Tertre, A., *et al.* (1992). The cholecystokinin hypothesis of panic and anxiety disorders: a review. *Journal of Psychopharmacology*, **6**, 345–51.
41. Ladouceur, R., Dugas, M.J., Freeston, M.H., *et al.* (1999) Specificity of generalized anxiety disorder symptoms and processes. *Behavior Therapy*, **30**, 191–207.
42. Weisberg, R.B., Brown, T.A., Wincze, J., *et al.* (2001). Causal attributions and male sexual arousal: the impact of attributions for a bogus erectile difficulty on sexual arousal, cognitions, and affect. *Journal of Abnormal Psychology*, **110**, 324–34.
43. Rapee, R.M. (1991). Generalized anxiety disorder: a review of clinical features and theoretical concepts. *Clinical Psychology Review*, **11**, 419–40.
44. Borkovec, T.D. (1994). The nature, functions, and origins of worry. In *Worrying: perspectives on theory, assessment, and treatment* (eds. G.C.L. Davey and F. Tallis). Wiley, New York.
45. Dugas, M.J., Gagnon, F., Ladouceur, R., *et al.* (1998) Generalized anxiety disorder: a preliminary test of a conceptual model. *Behavior Research Therapy*, **36**, 215–26.
46. Chorpita, B.F. and Barlow, D.H. (1998). The development of anxiety: the role of control in the early environment. *Psychological Bulletin*, **124**, 3–21.

47. Dumas, J.E., LaFreniere, P.J., and Serketich, W.J. (1995). 'Balance of power': a transactional analysis of control in mother-child dyads involving socially competent, aggressive, and anxious children. *Journal of Abnormal Psychology*, **104**, 104–13.
48. Silove, D., Parker, G., Hadzi-Pavlovic, D., *et al.* (1991). Parental representations of patients with panic disorder and generalized anxiety disorder. *The British Journal of Psychiatry*, **159**, 835–41.
49. Roemer, L., Molina, S., Litz., B.T., *et al.* (1996). Preliminary investigation of the role of previous exposure to potentially traumatizing events in generalized anxiety disorder. *Depression and Anxiety*, **4**, 134–8.
50. Torgersen, S. (1986). Childhood and family characteristics in panic and generalized anxiety disorders. *The American Journal of Psychiatry*, **143**, 630–2.
51. Steketee, G. and Foa, E.B. (1987). Rape victims: post traumatic stress responses and their treatment: a review of the literature. *Journal of Anxiety Disorders*, **1**, 69–86.
52. Ben-Noun, L. (1998). Generalized anxiety disorder in dysfunctional families. *Journal of Behavior Therapy and Experimental Psychiatry*, **29**, 115–22.
53. Yonkers, K.A., Warshaw, M.G., Massion, A.O., *et al.* (1996). Phenomenology and course of generalised anxiety disorder. *The British Journal of Psychiatry*, **168**, 308–13.
54. Yonkers, K.A., Bruce, S.E., Dyck, I.R., *et al.* (2003). Chronicity, relapse, and illness-course of panic disorder, social phobia, and generalized anxiety disorder: findings in men and women from 8 years of follow-up. *Depression and Anxiety*, **17**, 173–9.
55. Craske, M.G. and Barlow, D.H. (2006). *Mastery of your anxiety and worry* (2nd edn). Oxford University Press, New York.
56. Rickels, K., Downing, R., Schweizer, E., *et al.* (1993). Antidepressants for the treatment of generalized anxiety disorder: a placebo-controlled comparison of imipramine, trazodone, and diazepam. *Archives of General Psychiatry*, **50**, 884–95.
57. Rickels, K. (1990). Buspirone in clinical practice. *The Journal of Clinical Psychiatry*, **51**(Suppl. 9), 51–4.
58. Apter, J.T. and Allen, L.A. (1999). Buspirone: future directions. *Journal of Clinical Psychopharmacology*, **19**, 86–93.
59. Cutler, N.R., Sramek, J.J., Keppel-Hesselink, J.M., *et al.* (1993). A double-blind, placebo-controlled study comparing the efficacy and safety of ipsapirone versus lorazepam in patients with generalized anxiety disorder: a prospective multicenter trial. *Journal of Clinical Psychopharmacology*, **13**, 429–37.
60. Hedges, D.W., Reimherr, F.W., Strong, R.E., *et al.* (1996). An open trial of nefazodone in adult patients with generalized anxiety disorder. *Psychopharmacology Bulletin*, **32**, 671–6.
61. Rickels, K., Zanielli, R., McCafferty, J., *et al.* (2003). Paroxetine treatment of generalized anxiety disorder: a double-blind, placebo-controlled study. *The American Journal of Psychiatry*, **160**, 749–56.
62. Sheehan, D.V. (1999). Venlafaxine extended release (XR) in the treatment of generalized anxiety disorder. *The Journal of Clinical Psychiatry*, **60**(Suppl. 22), 23–8.
63. Davidson, J.R., DuPont, R.L., Hedges, D., *et al.* (1999). Efficacy, safety, and tolerability of venlafaxine extended release and buspirone in outpatients with generalized anxiety disorder. *The Journal of Clinical Psychiatry*, **60**, 528–35.
64. Barlow, D.H., Allen, L.B., and Basden, S.L. (2007). Psychosocial treatments for panic disorders, phobias, and generalized anxiety disorder. In *A guide to treatments that work* (3rd edn) (ed. P.E. Nathan and J.M. Gorman). Oxford University Press, New York.
65. Mitte, K. (2005). Meta-analysis of cognitive-behavioral treatments for generalized anxiety disorder: A comparison with pharmacotherapy. *Psychological Assessment*, **4**, 224–7.
66. Orsillo, S.M., Roemer, L., and Barlow, D.H. (2003). Integrating acceptance and mindfulness into existing cognitive-behavioral treatments for GAD. *Cognitive and Behavioral Practice*, **10**, 222–30.
67. Power, K.G., Simpson, R.J., Swanson, V., *et al.* (1990). A controlled comparison of cognitive-behaviour therapy, diazepam, and placebo, alone and in combination, for the treatment of generalized anxiety disorder. *Journal of Anxiety Disorders*, **4**, 267–92.
68. Swinson, R.P., Cox, B.J., and Fergus, K.D. (1993). Diagnostic criteria in generalized anxiety disorder treatment studies. *Journal of Clinical Psychopharmacology*, **13**, 455.
69. Sanderson, W.C., Beck, A.T., and McGinn, L.K. (1994). Cognitive therapy for generalized anxiety disorder: significance of comorbid personality disorders. *Journal of Cognitive Psychotherapy*, **8**, 13–18.
70. Bruce, S.E., Yonkers, K.A., Otto, M.W., *et al.* (2005). Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: a 12-year prospective study. *The American Journal of Psychiatry*, **162**, 1179–87.
71. Borkovec, T.D., Abel, J.L., and Newman, H. (1995). Effects of psychotherapy on comorbid conditions in generalized anxiety disorder. *Journal of Consulting and Clinical Psychology*, **63**, 479–83.
72. Boulenger, J., Fournier, M., Rosales, D., *et al.* (1997). Mixed anxiety and depression: from theory to practice. *The Journal of Clinical Psychiatry*, **58**(Suppl. 8), 27–34.

4.7.2 Social anxiety disorder and specific phobias

Michelle A. Blackmore, Brigette A. Erwin, Richard G. Heimberg, Leanne Magee, and David M. Fresco

Introduction

As our classification systems have been refined, we have come to view social anxiety disorder (social phobia) and specific phobias as distinct disorders, with divergent patterns of prevalence, aetiology, and course. Moreover, treatments for these disorders have become increasingly sophisticated. This chapter presents an overview of the current state of the field with regard to social anxiety disorder and specific phobias.

Social anxiety disorder

In the first and second editions of the *Diagnostic and Statistical Manual of Mental Disorders* (**DSM**),[1,2] all phobias were grouped together. However, in 1966 Marks and Gelder[3] observed that various phobias had different ages of onset and symptom presentations, providing the initial impetus for the inclusion of *social phobia* as a distinct disorder in DSM-III.[4] At first, research into the nature and treatment of social phobia lagged behind that of other anxiety disorders, leading to its description in 1985 as the neglected anxiety disorder.[5] Over the past two decades, however, attention to the conceptualization, definition, and classification of social phobia has increased dramatically. To acknowledge the significant impairment now known to be associated with social phobia and its differentiation from specific phobia, the alternative (and increasingly utilized) label, *social anxiety disorder*, was added in DSM-IV.[6]

Clinical presentation

Anxiety in situations involving potential evaluation by others (e.g. job interviews, public speaking engagements, first dates) falls within the realm of 'normal' social anxiety. For individuals with social anxiety disorder, however, situations such as these are typically associated with incapacitating levels of anxiety and a desire for escape or avoidance. Socially anxious individuals are often self-critical and perfectionistic and go to great lengths to avoid the negative evaluation of others that they may perceive as epidemic. Commonly, persons with social anxiety disorder experience somatic symptoms such as blushing, trembling, dry mouth, or perspiring, which they believe will be noticed by others and provide further evidence of their incompetence. Children may manifest their anxiety differently than adults; they may cry, throw tantrums, freeze, shrink from interactions with strangers, and they may not acknowledge that their fears are irrational.[6] By leaving anxiety-provoking situations (escape) or foregoing them entirely (avoidance), individuals with social anxiety disorder may reduce or prevent the immediate experience of anxiety, but this behaviour reinforces beliefs in their inadequacies and serves to maintain anxiety in the absence of objective threat.[7,8]

Functional impairment

Individuals with social anxiety disorder experience significant impairment in social, educational, and occupational functioning.[5,9] They are less likely to marry and are more likely to divorce than those without the disorder.[10] They also have fewer friends and more trouble getting along with the friends they have than persons without the disorder.[11] Individuals with social anxiety disorder assessed in a primary care setting reported missing an average of 3 days of work and having an average of 6 days of reduced productivity in the last month because of their emotional problems.[12] Comparatively, mentally healthy individuals reported less than 1 day of lost work and reduced productivity combined. Unemployment, underemployment (working at a level below the individual's abilities), and financial dependency are also characteristic of individuals with social anxiety disorder.[10]

Classification

(a) DSM and ICD

Whereas the DSM is widely used in North America, the *International Classification of Mental and Behavioural Disorders* (**ICD**) is commonly used in other parts of the world. Social anxiety disorder, termed social phobia in ICD-10,[13] first appeared in ICD-10 12 years after its appearance in DSM-III. The ICD-10 criteria for social phobia are less detailed and more circumscribed than those in DSM-IV. Specifically, DSM-IV requires excessive fear of humiliation or embarrassment in social or performance situations, anxiety provoked by exposure to feared situations, recognition that the fear is excessive, avoidance of situations or endurance with distress, and significant distress or impairment. Further, the fear and avoidance cannot be better accounted for by another psychiatric disorder, a general medical condition, or the effects of a substance. The ICD-10, in contrast, requires only that the symptoms be representative of anxiety and not another psychiatric disorder, that the anxiety occurs in relation to social situations, and that avoidance of anxiety-provoking situations be present. Because most published research on social anxiety disorder relies on DSM rather than ICD criteria, this chapter will do so also.

(b) Diagnostic issues

Individuals presenting for treatment of social anxiety disorder endorse multiple fears and significant impairment. The *generalized* subtype is specified when 'most social situations' are feared, whereas the *non-generalized* subtype describes persons who fear a more limited set of social situations. Individuals with generalized and non-generalized social anxiety disorder differ on several dimensions, including symptom severity, functional impairment, and physiological symptoms when exposed to feared situations.[14] Conclusive differences between subtypes in course and response to treatment remain to be demonstrated.[15,16]

Like social anxiety disorder, avoidant personality disorder is regarded as an extreme fear of negative evaluation, leading researchers to view the two conditions on a continuum that is artificially divided at the boundary between Axes I and II. Many investigators conclude that the co-occurrence of generalized social anxiety disorder and avoidant personality disorder represent persons with the most severe social anxiety and the poorest global functioning.[17]

Social anxiety may also develop as a result of medical conditions, such as becoming excessively anxious or avoiding social situations because of obesity, acne, benign essential tremor, stuttering, or the disability associated with Parkinson's disease. These conditions are not considered exemplars of social anxiety disorder because anxiety developed secondary to the medical condition. Rather, they are assigned to the category 'anxiety disorder not otherwise specified'. However, persons who experience secondary social anxiety are often responsive to pharmacological or cognitive behavioural treatments with demonstrated efficacy for social anxiety disorder.[18]

(c) Comorbidity and differential diagnosis

Approximately 81 per cent of persons with primary social anxiety disorder meet criteria for at least one other lifetime psychiatric disorder.[19] Social anxiety disorder most commonly co-occurs with other anxiety disorders,[20] although comorbid diagnoses of depression and alcohol use disorders are also common. Differential diagnosis is complicated by the fact that certain Axis I disorders both resemble and co-occur with social anxiety disorder.

(d) Social anxiety disorder versus panic disorder with agoraphobia (PDA)

PDA can be differentiated from social anxiety disorder in several ways. Although many individuals with social anxiety disorder experience panic attacks, the attacks occur in anticipation of negative evaluation by others. For persons with panic disorder, panic attacks are often unexpected, may not be associated with specific cognitions, and can be nocturnal.[5] Persons with social anxiety disorder are more likely to experience blushing and muscle twitches, whereas individuals with PDA are more likely to experience symptoms such as blurred vision, headaches, chest pain, ringing in the ears, and fear that they will die or go crazy.[21] The age of onset for social anxiety disorder tends to be earlier than that for PDA.[22] Individuals presenting for social anxiety disorder treatment either show an equal gender distribution or are slightly more likely to be male,[23] whereas those presenting for PDA treatment are substantially more likely to be female.[21] Finally, persons with social

anxiety disorder report feeling more comfortable when alone, whereas persons with PDA may feel more at ease in the presence of others.[22]

(e) Social anxiety disorder versus generalized anxiety disorder (GAD)

Individuals with GAD endorse higher levels of social anxiety than other persons with non-social anxiety disorders.[23] Although individuals with either social anxiety disorder or GAD may devote excessive amounts of time to worrying and ruminating, the focus of worry in social anxiety disorder is on fear of evaluation in social or performance situations, whereas the hallmark feature of worry in GAD is heightened focus on possible catastrophic consequences across several domains of life. Persons with social anxiety disorder are more likely to experience sweating, flushing, and breathing problems; those with GAD more commonly experience headaches, insomnia, and fear of dying.[24]

(f) Social anxiety disorder versus depression

Social anxiety disorder and depression may have withdrawal from social situations in common.[21] In differentiating between the two disorders, one must consider the reason for this withdrawal. Persons with depression withdraw because they fail to experience pleasure or lack the energy for social engagement. Individuals with social anxiety disorder fear the negative evaluation they believe to be associated with such situations. Persons with depression may be indifferent about engaging in social situations, whereas individuals with social anxiety disorder often have a strong desire to affiliate with others that is hampered by anxiety.

Epidemiology

(a) Prevalence

The National Comorbidity Survey Replication (**NCS-R**) reported a lifetime prevalence rate of 12.1 per cent[25] and a 12-month prevalence rate of 6.8 per cent[20] for social anxiety disorder. NCS-R lifetime prevalence rates render social anxiety disorder the fourth most common psychiatric disorder behind major depression (16.6 per cent), alcohol abuse (13.2 per cent), and specific phobia (12.5 per cent).

(b) Age at onset/age of treatment seeking

Social anxiety disorder often begins early in life. Mean age of onset for the disorder ranges from 13 to 20 years old, although patients often report having experienced symptoms for as long as they can remember.[26] Despite early onset, persons with social anxiety disorder often do not seek treatment for approximately 16 years after onset,[27] and many never do.[28]

(c) Gender differences

Although men and women with social anxiety disorder who seek treatment do so in relatively equal numbers,[23] epidemiological studies suggest that women are more likely than men to have the disorder.[19,25] In a clinical sample, women reported fear of more social situations and scored higher on several social anxiety disorder assessment measures.[29] Thus, although women are more likely to experience social anxiety, men are more likely to seek treatment and may do so when troubled with less severe symptoms. It may be that social anxiety disorder impairs the expected role functioning of men to a greater extent that it does for women.[29]

Aetiology of social anxiety disorder

Genetic factors appear to contribute to the emergence of social anxiety disorder. Higher rates of social anxiety disorder have been found in relatives of individuals with the disorder compared to relatives of persons without the disorder.[30] Further, rates of social anxiety disorder in first-degree relatives of probands with the generalized subtype are higher than in relatives of probands with the non-generalized subtype or with no psychiatric history.[31] Kendler *et al.*[32] report concordance rates for social anxiety disorder among monozygotic twins (24.4 per cent) to be greater than the rates for dizygotic twins (15.3 per cent). A study conducted with the same cohort 8 years later found the heritability of social anxiety disorder to be approximately 50 per cent in female twins[33] and 25 per cent in male twins.[34]

Neurobiological factors may also be associated with social anxiety disorder. Imaging studies of individuals with social anxiety disorder have demonstrated increased activity in regions associated with fear and anxiety (i.e. prefrontal cortex, amygdala, hippocampus) during anxiety-provoking tasks.[35] Given the efficacy of serotonin reuptake inhibitors and monoamine oxidase inhibitors in treating social anxiety disorder,[36] dysregulation of the serotonin[37] and dopamine[38] systems have been investigated as potential correlates of the disorder.

Several studies also suggest the importance of parental influences and significant life events in the development of social anxiety disorder. Persons with social anxiety recall observing their mothers act more fearful and avoidant of social interactions[39] and describe their parents as overprotective.[40] Stressful social and performance situations early in life (e.g. public ridicule, being bullied, mind going blank during a presentation) are also commonly reported by persons with social anxiety disorder.[41]

Course of social anxiety disorder

Social anxiety disorder is chronic and unlikely to remit without treatment. The disorder persists throughout adulthood[42] and its course is unrelated to gender, age of onset, duration of illness, level of functioning at intake, lifetime history of anxiety disorders, or current comorbidity of anxiety or depressive disorders.[42,43] Two conditions related to social anxiety disorder emerge in childhood and are relatively stable into adulthood—shyness and behavioural inhibition. Individuals who had been shy as children exhibited overall lower levels of functioning when assessed 30 years later.[44] Similarly, children described as behaviourally inhibited, or having the tendency to withdraw from novel people, settings, or objects, have demonstrated increased risk for the development of social anxiety disorder in adolescence.[45] Behavioural inhibition was also more prevalent in children of individuals with anxiety disorders and remained relatively stable for over 7 years in children initially assessed between the ages of 21 to 31 months.[46] These findings suggest that extreme shyness and behavioural inhibition may be early manifestations of social anxiety disorder.

Empirically evaluated treatments

(a) Cognitive behavioural interventions

Cognitive behavioural treatments have been subjected to the most thorough evaluation in the empirical literature. Treatments that utilize exposure alone or combined with cognitive restructuring have received the greatest empirical support and are the focus of

our review. Because of space limitations, other cognitive behavioural treatments, such as social skills training and applied relaxation, will not be reviewed, and the reader is referred to other sources.[47]

(b) Exposure

Exposure requires individuals to imagine (imaginal exposure) or directly confront (*in vivo* exposure) feared stimuli. Research examining the efficacy of imaginal exposure for social anxiety disorder is limited; however, *in vivo* exposure has repeatedly demonstrated short- and long-term efficacy in therapist-directed and self-directed formats.[47] Exposure requires patients to progressively confront anxiety-provoking situations beginning with situations that elicit moderate fear. Patients turn to the next most feared situation after repeated and prolonged exposure to the previous situation no longer elicits a distressing level of fear. Individuals with social anxiety disorder treated with exposure alone experienced greater improvement than individuals receiving relaxation training,[48] pill placebo,[49] or delayed treatment.[50]

(c) Exposure combined with cognitive restructuring

Contemporary cognitive behavioural models of social anxiety disorder propose that anxiety is largely maintained by dysfunctional beliefs and information-processing biases, and that successful treatment will be associated with modification of cognition.[7,8] Accordingly, exposure is typically combined with techniques designed to modify dysfunctional thinking patterns.[51,52]

Cognitive restructuring is an intervention based on the theory that one's thoughts about a situation, not the situation itself, generate anxiety.[53] The intervention is designed to help patients challenge maladaptive beliefs by identifying irrational thoughts, evaluating the dysfunction inherent in these thoughts, and deriving rational alternatives to these thoughts. By engaging in this process and utilizing alternative thoughts during exposure to feared situations, patients acquire new, adaptive learning that competes with their previously-learned maladaptive views, and, in turn, lessens the anxiety they experience.[54]

Efficacy for the combination of cognitive restructuring and exposure has been demonstrated in comparison to wait-list control conditions,[55,56] pill placebo,[57] and psychological placebo conditions.[57,58] Several studies demonstrate equivalent outcomes for exposure alone and exposure plus cognitive restructuring, whereas others indicate the combination shows superior efficacy and additional gains at follow-up.[47] In one meta-analysis,[59] only the combination of exposure and cognitive restructuring was superior to placebo treatments, but this difference has not been reliably demonstrated.[60] Nevertheless, patients treated with exposure alone tend to show deterioration of gains after treatment, suggesting durability of gains may be enhanced with the addition of cognitive restructuring techniques.

(d) Cognitive behavioural group therapy (CBGT)

CBGT, originally developed by Heimberg and Becker,[51] is one of the most thoroughly examined cognitive behavioural treatments for social anxiety disorder. It integrates cognitive techniques and exposure and is typically conducted in 12 weekly, 2.5 h sessions, with approximately six patients and two therapists. In sessions 1–2, patients receive psychoeducation, rationale and instructions for exposure, training in cognitive restructuring, and homework assignments. Thereafter, therapists lead patients through individualized exposures preceded and followed by therapist-directed cognitive restructuring exercises. For homework, patients practice cognitive restructuring before and after exposure to real-life anxiety-provoking situations.

One study evaluating the efficacy of CBGT compared it to educational-supportive group therapy (ES), a credible placebo treatment consisting of lectures, discussions, and social support. Seventy-five per cent of CBGT patients made significant improvement compared to 40 per cent of ES patients.[58] At follow-up (4.5–6.25 years), CBGT produced durable treatment gains.[61] A comparison of CBGT to the monoamine oxidase inhibitor phenelzine, pill placebo, and ES demonstrated equivalent response and attrition rates after 12 weeks of treatment for CBGT and phenelzine, both of which were superior to placebo and ES.[57] Although the phenelzine group evidenced superior improvement on a subset of measures after 12 weeks, CBGT demonstrated more durable treatment gains, with only 17 per cent relapse compared to 50 per cent relapse in the phenelzine group after a 6-month maintenance phase and 6-month follow-up phase.[62] An intensive version of CBGT, based on a hybrid of the treatment approaches developed by Heimberg and by Clark, involving 2 weeks of daily treatment sessions separated by 1 week of homework assignments, also proved superior to a wait-list control.[63]

(e) Individual cognitive behavioural therapy

Group CBT may not always be feasible, particularly in clinical settings where it may be difficult to obtain an adequate number of patients to form a group. However, CBGT has been adapted to an individual format and proven superior to a wait-list control (Heimberg, unpublished observations). Clark[52] also developed a cognitively-focused individual treatment for social anxiety disorder that has demonstrated substantial efficacy. The treatment instructs patients on how to shift their attention externally (rather than on the self) and to reduce reliance on safety behaviours. Video feedback and exposure to feared situations aimed at restructuring distorted cognitions are also incorporated.

Individual cognitive therapy demonstrated superior efficacy to wait-list control, with clinically significant gains observed in 76 per cent of patients receiving cognitive therapy, compared to 38 per cent of patients receiving an applied relaxation treatment and 0 per cent of patients in the wait-list control group.[64] Cognitive therapy was also more efficacious than fluoxetine plus self-exposure instructions and placebo plus self-exposure instructions, with gains maintained at 1-year follow-up.[65] Although meta-analytic studies suggest that individual and group CBT are similar in efficacy,[60] individual cognitive therapy was superior to a group therapy based on Clark's model on several measures.[66] Similarly, individual cognitive therapy proved superior to a 3-week intensive group therapy based on Clark's model and treatment with psychiatrist-selected SSRIs.[67]

Pharmacotherapy

The efficacy of selective serotonin uptake inhibitors (**SSRI**s), serotonin–norepinephrine reuptake inhibitor venlafaxine (**SNRI**s), benzodiazepines, reversible and irreversible monoamine oxidase inhibitors (**MAOI**s), ß-adrenergic blockers, buspirone, gabapentin, and pregabalin have been evaluated in placebo-controlled studies. To date, most controlled trials have included patients with generalized social anxiety disorder, so it is unclear if similar efficacy would be observed among patients with the non-generalized subtype of the disorder.

(a) Selective serotonin reuptake inhibitors and venlafaxine

The SSRIs (e.g. paroxetine, sertraline, fluoxetine, fluvoxamine, and escitalopram) and the SNRI venlafaxine are regarded as first-line medication treatments for generalized social anxiety disorder, given demonstrated efficacy in placebo-controlled trials for the disorder as well as for other anxiety and affective disorders. Further, potential adverse side effects of SSRIs (e.g. sweating, sexual dysfunction) pose a lower risk than potential adverse side effects of the MAOIs (e.g. high blood pressure, seizure) or benzodiazepines (e.g. dependence, overdose).

The SSRIs have demonstrated superior efficacy to placebo as well as CBT in the short-term,[68] with the possible exception of fluoxetine.[65,69] There have been few direct comparisons of the various SSRIs or SNRIs. One study comparing escitalopram (5, 10, 20 mg), paroxetine (20 mg), and placebo, demonstrated superior efficacy for 5 mg and 20 mg of escitalopram and 20 mg of paroxetine compared to placebo at week 12.[70] At week 24, all doses of escitalopram proved superior to placebo and 20 mg of escitalopram proved superior to paroxetine on a subset of social anxiety measures. The SNRI venlafaxine produced results comparable to paroxetine, including superior efficacy relative to placebo.[71]

A fixed-dose trial of paroxetine showed 20 mg/day to be an efficacious and safe dose for the treatment of social anxiety disorder, with superior efficacy to placebo.[72] In a flexible-dose trial, paroxetine proved more efficacious than placebo after 8–12 weeks of treatment, with 55 per cent of paroxetine patients compared to 24 per cent of placebo patients identified as much or very much improved.[73] A double-blind discontinuation study demonstrated a 14 per cent relapse rate in the paroxetine group compared to a 39 per cent relapse rate in the placebo group at week 24.[74]

Van Ameringen and colleagues[75] found sertraline to be more efficacious than placebo, and a 20-week double-blind discontinuation trial demonstrated significantly lower relapse rates for patients continuing on sertraline than patients switched to placebo (4 per cent versus 36 per cent, respectively).[76] A study of fluvoxamine demonstrated superior outcome, beginning at week 8, for patients receiving fluoxamine (20 mg mean daily dose) than patients receiving placebo.[77]

Fluoxetine has yielded less promising results and is the only SSRI to date that has failed to separate from placebo in controlled trials.[65,69] However, in one study fluoxetine evidenced superior efficacy to placebo, although no differences were found between fluoxetine, CBGT, or their combination.[78]

(b) Benzodiazepines

Research suggests benzodiazapines have performed better than MAOIs and CBT in the short-term[68] and offer rapid onset of effect, good tolerability, and minimal side effects (e.g. sedation, withdrawal effects with abrupt discontinuation). However, benzodiazepines are contraindicated in the presence of substance abuse and lack efficacy for comorbid mood disorders. Clonazepam was shown to be efficacious for social anxiety disorder (2–4 mg/day), with patients showing improvement after 6 weeks of treatment and superior response rates compared to placebo patients at week 10.[79] Clonazepam also produced equivalent gains to CBGT after 12 weeks of treatment.[80] One study of alprazolam demonstrated that the medication was no more efficacious than placebo.[81]

(c) Monoamine oxidase inhibitors

MAOIs have demonstrated efficacy in the treatment of social anxiety disorder[36] similar to that of SSRIs. However, MAOI therapy requires strict dietary restrictions as ingesting foods rich in tyramine (e.g. most aged cheeses, red wine, beer) can increase the risk of rapidly escalating blood pressure. MAOIs are also associated with more adverse side effects (e.g. hypotension, sedation, weight gain, sexual dysfunction) than SSRIs. However, MAOI therapy (usually 45–90 mg/day) is justified when other treatments with more benign side effect profiles prove ineffective.

As discussed earlier, phenelzine proved superior to placebo and demonstrated equivalent response rates to CBGT after 12 weeks of treatment,[57], although CBGT evidenced more durable treatment gains.[62] Liebowitz and colleagues[82] found that after 8 weeks of acute treatment and 8 weeks of maintenance, phenelzine produced results superior to those of placebo and the ß-adreneregic blocker atenolol. Versiani and colleagues[83] found patients who received phenelzine endorsed significantly greater reductions in social anxiety than placebo patients at 8 weeks. In a comparison of phenelzine, alprazolam, CBGT, and placebo with instructions for self-exposure,[81] all groups produced relatively equivalent response. However, phenelzine demonstrated maintenance of gains at a 2-month follow-up, whereas alprazolam did not.

(d) Reversible inhibitors of monoamine oxidase

Moclobemide and brofaramine have been evaluated for the treatment of social anxiety disorder and offer fewer side effects and dietary restrictions than MAOIs. However, moclobemide has shown modest efficacy in controlled trials and is less |efficacious than the SSRIs and MAOIs.[36] Although brofaramine was found to be efficacious for social anxiety, it was never marketed.

(e) ß-adreneregic blockers

ß-adreneregic blockers (e.g. atenolol) have failed to surpass placebo in controlled trials.[49] Propranolol appears to be efficacious on an as-needed basis (10–40 mg) for anxiety related to performance situations.

(f) Buspirone

Buspirone appears no more efficacious than placebo at low dosages.[84] One study demonstrated greater efficacy of buspirone in doses greater than or equal to 45 mg/day or when used as an augmentation to an SSRI.[85]

(g) Gabapentin and pregabalin

A single study shows some efficacy of high doses of the anticonvulsant gabapentin (maximum of 3600 mg/day) in the treatment of social anxiety disorder relative to placebo.[86] Pregabalin was also superior to placebo in one study,[87] but a high dose was required as well.

(h) Other agents

D-cycloserine, a partial adrenergic agonist associated with the facilitation of learning and memory consolidation, appears to enhance the efficacy of exposure. Several animal trials demonstrated enhanced fear extinction when d-cycloserine was administered prior to and after exposure trials.[88] Similar results were shown when individuals receiving d-cycloserine 1 h prior to exposure demonstrated significantly less social anxiety posttreatment than those receiving placebo.[89]

Integrating pharmacotherapy and cognitive behavioural interventions

Several studies have examined the utility of combining pharmacotherapy with cognitive behavioural interventions. Blomhoff and colleagues[90] compared sertraline and pill placebo alone and in combination with physician-directed exposure or general medical care, which included non-directive encouragement. After 12 weeks, all active treatments surpassed pill placebo and non-directive encouragement, with sertraline and exposure proving equally efficacious. Only sertraline alone or in combination with exposure proved superior to placebo at 24 weeks. However, only patients who received exposure alone demonstrated further improvement at 1-year follow-up, whereas patients receiving sertraline with or without exposure showed some degree of deterioration.[91]

Preliminary data from a study comparing CBGT, phenelzine, their combination, and pill placebo, suggests phenelzine plus CBGT may be more likely to surpass placebo than either treatment alone.[92] A similar study by Davidson and colleagues[78] found no advantage in combining fluoxetine and CBGT. Preliminary data suggests the combination of d-cycloserine with exposure may enhance the efficacy of exposure treatment.[89]

Evidence for the utility of the combination of these treatments is mixed. To summarize, different studies suggest that combining some medication treatments with CBT is no more efficacious than either treatment alone;[78] combining medication with CBT may diminish the efficacy of CBT;[91] and combining treatments may show only modest benefits over the administration of either alone.[92] The hypothesis that sequential treatments might capitalize on rapid medication response for faster symptom relief and cognitive behavioural treatment for superior relapse protection needs evaluation. Cognitive behavioural treatments may be utilized to help patients discontinue medical regimens after initial drug response.

Management of social anxiety disorder

(a) Treatment selection

Given the chronic nature of social anxiety disorder, CBT offers more enduring treatment gains compared to medication and lacks the adverse side effects. When considering cognitive behavioural interventions, it is important to assess the patient's ability and willingness to endure exposure to anxiety-provoking situations, as exposure necessitates short-term increases in anxiety over and above what the individual may typically experience. Additionally, it is helpful to assess level of compliance with previous therapy experiences (e.g. homework compliance). Managing a patient's treatment expectancy may further improve outcome as individuals reporting higher expectancy regarding the efficacy of treatment have demonstrated greater improvement and maintenance of treatment gains.[93]

Medication may be a preferred choice for individuals requiring rapid response, CBT non-responders, those with a preference for medication, or those with severe social anxiety disorder and/or comorbid depression. When selecting medication treatment, it is important to consider possible side effects, the patient's physical health, and prior medication compliance. Benzodiazepines offer rapid onset of effect, particularly beneficial for short-term performance situations, but their use may be contraindicated among patients with a history of alcohol or substance abuse. Given these concerns and their more benign side effect profile, SSRIs and venlafaxine are best regarded as the first-line medications, with the benzodiazepines and MAOIs held in reserve for non-responding patients. To date, we know little of the appropriate duration of medication treatments in social anxiety disorder, although two studies evidenced significant relapse rates upon discontinuation of SSRIs after a total treatment period of 36–40 weeks.[74,76]

(b) Other issues in management

Comorbid conditions should also be considered when treating social anxiety disorder with cognitive behavioural or pharmacological treatments. Comorbid anxiety disorders do not appear to degrade a patient's response to CBT; however, patients with comorbid mood disorders may present with more severe social anxiety symptoms and, although they improve at a similar rate, require extended treatment.[94] In general, if any comorbid condition prevents the patient from engaging in exposure exercises, taking medications in the prescribed manner, or complying with any other therapeutic activities, it may be necessary to treat the comorbid condition before the social anxiety disorder. In other cases, a comorbid condition might become the focus after social anxiety disorder treatment. Patients who abuse substances to 'treat' their social anxieties are often reticent to give up their substances before other coping strategies are made available. Ultimately, the management of comorbid conditions in individuals with social anxiety disorder depends on the analysis of the relationships between the disorders and varies from case to case.

(c) Prevention

Few studies to date have specifically examined ways of preventing social anxiety disorder; however, evidence for familial aggregation and environmental influence is strong. Parents may reinforce anxious children for making avoidant choices.[95] Thus, it is important to consider including parents of socially anxious children in treatment to provide parents with strategies to help their child manage the social anxiety. Parents with social anxiety disorder may also benefit from treatment themselves.

Since social anxiety disorder has an early age of onset, treatment of children and adolescents should help prevent social anxiety from becoming a chronic condition. Prevention programmes have been integrated into school settings to teach children and parents coping skills (e.g. cognitive restructuring, relaxation) and instruct them on the proper conduct of exposures. Individuals in the intervention group have shown significant improvement compared to non-intervention controls.[96] Similar prevention programmes have targeted at-risk youth and resulted in decreased anxiety symptoms and lower rates of anxiety disorders compared to control groups.[97]

Informational Websites

- http://www.adaa.org/GettingHelp/AnxietyDisorders/SocialPhobia.asp (Anxiety Disorders Association of America)
- http://www.nimh.nih.gov/HealthInformation/socialphobia-menu.cfm (National Institute of Mental Health)

Specific phobia

Clinical features and functional impairment

The hallmark feature of specific phobia, prior to DSM-IV called simple phobia, is a 'marked and persistent fear that is excessive or

unreasonable, cued by the presence or anticipation of a specific object or situation' (6, p. 405). Commonly feared/avoided objects include animals, aspects of nature, or blood. Many individuals endorse some fear of these stimuli; however, in specific phobia, fear and avoidance cause significant interference with one's normal routine, career, academic pursuits, or social activities. Some individuals with specific phobias maintain a relatively normal routine by pursuing a lifestyle that minimizes exposure to the phobic stimulus. Often, specific phobias accompany a more impairing primary disorder that is the stimulus for seeking treatment.[98]

Classification

(a) DSM and ICD

Criteria for specific phobia are similar between DSM-IV and ICD-10. Both view fear as arising from exposure to a specific object or situation, which leads to acute autonomic and psychological symptoms of anxiety. However, ICD-10 does not stipulate that anxiety may be cued by the anticipation of the feared object or situation. Our review of specific phobia follows DSM conventions.

(b) Specific phobia subtypes

DSM-IV classifies specific phobias into five subtypes:

- animal
- natural environment
- blood-injection-injury
- situational
- other (e.g. dental/medical procedures, choking, etc.).

With the exception of blood-injection-injury phobias, exposure to the phobic stimulus evokes intense anxiety that may meet criteria for a situationally-bound panic attack. Additionally, there is extreme apprehension and desire to escape or avoid the phobic stimulus.[99] By contrast, individuals with blood-injection-injury phobias exhibit a biphasic anxiety reaction (vasovagal syncope) characterized by initial, short-lived sympathetic arousal, followed by parasympathetic arousal that may result in fainting.[100] The subjective experiences of these individuals tend to be characterized by disgust and repulsion rather than pure apprehension.[99]

(c) Comorbidity

The vast majority (83.4 per cent) of individuals with specific phobia experience at least one other lifetime psychiatric disorder.[19] In the original National Comorbidity Survey, individuals with specific phobia were 5 times more likely to have at least one additional disorder than individuals who had never met criteria for a phobic disorder. In most cases, the onset of the specific phobia preceded the onset of the other disorder.

Epidemiology

(a) Prevalence

The NCS-R reported a lifetime prevalence rate of 12.5 per cent[25] and a 12-month prevalence rate of 8.7 per cent for specific phobia.[20] One study found situational/environmental phobias to be the most common (13.2 per cent), followed by animal phobias (7.9 per cent) and blood-injection-injury phobias (3.0 per cent).[101]

(b) Age at onset/age of treatment seeking

Age at onset of specific phobia tends to be earlier than other anxiety disorders[19,102] and varies as a function of phobia subtype. For example, animal phobias onset earliest (age 7), followed by blood-injection-injury phobias (age 8), and most situational phobias (early 20s).[103] Individuals with phobic disorders often do not seek treatment for 20 years after onset.[27]

(c) Gender distribution

Women receive diagnoses of specific phobia more often than men. Lifetime rates for specific phobia in the NCS-R were 15.7 per cent for women but only 6.7 per cent for men.[25] In one study, women reported higher rates of animal and situational/environmental phobias, but rates of blood-injection-injury phobia did not differ.[101]

Aetiology of specific phobia

There is considerable evidence for familial/genetic transmission of specific phobia.[32,104] In one study, 31 per cent of first-degree relatives of persons with specific phobia also met criteria for specific phobia.[104] Rates of specific phobia were higher among first-degree relatives of persons with specific phobia and no other anxiety disorder than among first-degree relatives of persons who were never mentally ill.[105] In the Virginia Twin Study,[32] concordance rates for animal phobia were 25.9 and 11.0 per cent among monozygotic and dizygotic twins, respectively. Concordance rates for situational phobia were similar in monozygotic and dizygotic twins. Moreover, children classified as behaviourally inhibited have shown higher risk for the development of multiple specific phobias.[106]

Classical conditioning theory[107] holds that phobias are learned through the association of negative experience with an object or situation. However, individuals with more previous non-traumatic experiences with the object or situation are less likely to develop a phobia upon a negative encounter than those with less prior experience when traumatized. Two-factor learning theory[108] introduced avoidance as a critical component to the maintenance of anxiety. That is, responses of avoidance or escape are learned and serve to decrease the discomfort arising from exposure to conditioned stimuli. Repeated negative reinforcement of avoidance behaviour (i.e. reduction of arousal on removal of oneself from proximity to the phobic object or situation) maintains the fear and makes it resistant to extinction.

Some conditioning theorists assert that feared stimuli are not randomly determined; rather, humans have inherited a predisposition to fear specific stimuli through natural selection. Marks'[109] 'preparedness' theory maintains that commonly feared objects are those that historically threatened the survival of the individual or the species. In this model, phobias are viewed as instances of 'prepared learning', which is selective, easily acquired, difficult to extinguish, and non-cognitive.[110] However, a large number of studies also suggest that phobias may be acquired via observational and informational learning (e.g. hearing that the situation is dangerous).

Course of specific phobia

Individuals with specific phobias acquire their fear(s) early in life, and the disorder tends to be chronic or recurrent without treatment.[111] Often individuals with specific phobias adapt their lifestyle to avoid contact with the feared stimuli, such that only

persons with the most severe specific phobias seek treatment. Events that commonly precipitate treatment seeking include a change in lifestyle that increases exposure to the feared stimulus (e.g. change in occupation that requires frequent air travel) or the experience of a panic attack in anticipation or in the presence of the feared stimulus.

Empirically evaluated treatments

(a) Cognitive behavioural interventions

(i) Exposure

Prolonged and repeated *in vivo* exposure to feared stimuli is by far the most studied and efficacious intervention for specific phobia[112] and should be considered the first-line treatment. Although *in vivo* exposure is generally believed to be more efficacious than imaginal exposure for specific phobia, some studies found *in vivo* and imaginal exposure techniques to be similarly efficacious.[113] Modelling in the form of observing another patient receive treatment has been shown to enhance the effects of exposure and to increase the speed at which positive outcomes are attained.[114] Multiple exposures sessions are considered more efficacious than one session, although some studies report positive outcomes for one-session treatments.[115] Further, one-session group *in vivo* exposure treatment has produced gains similar to those of one-session individual *in vivo* exposure treatment.[115] Variations in spacing of exposure sessions (e.g. 10 daily versus 10 weekly) have shown equivalent outcomes.[116] Finally, therapist-directed treatments have generally been more efficacious than self-directed treatments,[117] with gains enduring up to 8 years.[114] When possible, exposures should be conducted in a variety of settings to enhance generalization outside the therapeutic setting.[54]

In vivo exposure situations can sometimes be difficult to arrange and imaginal exposure may not achieve the reality or intensity needed to elicit an anxiety response. In such instances, virtual reality exposure (**VRE**), in which feared situations are presented in three dimensional simulations, may greatly enhance the efficacy of exposures. The salience of virtual environments can be augmented by instructing patients to touch real objects (e.g. toy spiders) which correspond with the virtual environment. Recent case studies demonstrated the efficacy of VRE either alone[118,119] or in combination with anxiety management training.[120] In a controlled trial for fear of flying, VRE training showed equivalent gains to *in vivo* exposure at 6 and 12-month follow-up[121]. Another study indicated that VRE produced superior gains to an attention placebo therapy, although gains were not maintained at 6-month follow-up.[122] D-cycloserine in combination with VRE has also shown enhanced treatment efficacy over VRE with placebo.[123]

(b) Applied relaxation and applied tension

Applied relaxation combines focused attention on different muscle groups while tensing and relaxing muscles,[124] with instruction to practice these skills first in non-anxiety-provoking situations and then in anxiety-provoking situations.[125] Although research is limited, applied relaxation has demonstrated efficacy,[126] especially among patients with higher levels of physiological reactivity than behavioural avoidance. *Applied tension*, designed specifically for blood-injection-injury phobia to treat parasympathetic arousal, requires the patient to tense different muscle groups in the presence of phobic stimuli to elevate blood pressure. Persons with phobias for blood, wounds, and injuries responded equally well to applied tension, applied relaxation, or their combination[127] Individuals treated with applied tension also evidenced greater treatment gains posttreatment and at 1-year follow-up than those treated with *in vivo* exposure alone.[128] In one dismantling study, individuals treated with applied tension and tension only (without the exposure component) evidenced equivalent gains and maintained superior outcomes at posttreatment and 1-year follow-up compared to patients who received *in vivo* exposure.[129]

(c) Cognitive restructuring

Phobia-specific irrational thoughts may contribute to the development of the phobia, maintain avoidance behaviour, and contribute to physiological symptoms.[130] When combined with exposure to feared stimuli, cognitive restructuring has proven efficacious,[131] although there are relatively few studies of cognitively-oriented treatments for specific phobia.

Pharmacotherapy

Drug treatments for specific phobia have consistently been shown to be less efficacious than behavioural treatments and may hinder maintenance of treatment gains. ß-Adrenergic blockers reduce some symptoms of sympathetic arousal during exposure to feared stimuli but fail to decrease subjective fear.[132] Although benzodiazepines (e.g. diazepam) may facilitate approach to feared stimuli, they may also reduce the efficacy of behaviour therapies by inhibiting the experience of anxiety during exposure.[133] Recent studies suggest d-cycloserine may facilitate exposure and extinction to specific feared stimuli. In a study of acrophobia patients, d-cycloserine was administered prior to exposure sessions and proved superior to placebo in reducing anxiety symptoms.[123]

Management of specific phobia

(a) Treatment selection

Exposure is clearly the treatment of choice for specific phobia, although facing feared stimuli may be particularly challenging for some patients, and their willingness and ability to participate in exposures should be assessed prior to treatment. Tailoring treatment to individual response patterns may improve outcome. For instance, patients with heightened physiological reactivity may respond preferentially to applied relaxation, whereas patients showing avoidance behaviour may respond better to *in vivo* exposure.[126] Further, individuals who experience anxiety primarily in the form of anxious thoughts may respond better to cognitive techniques.[134]

For patients unwilling or unable to engage in cognitive behavioural interventions, medication may be an appropriate alternative. However, it is first important to educate patients about the possibility of dependence with regular use and the side effects of high doses (e.g. sedation). If the patient is participating in exposure therapy, it is also necessary to explain that medication may interfere with the efficacy of exposure treatment.

(b) Prevention

Children with specific phobia were included in the Queensland Early Intervention and Prevention of Anxiety Project, but no other preventive efforts have been mounted. It is tempting to speculate that children could be 'inoculated' against a variety of the more common specific phobias by systematic exposure to potentially feared objects or situations.

Further information

http://www.adaa.org/GettingHelp/AnxietyDisorders/SpecificPhobia.asp (Anxiety Disorders Association of America)
http://www.mentalhealthamerica.net/go/phobias (Mental Health America)

References

1. American Psychiatric Association. (1952). *Diagnostic and statistical manual of mental disorders* (1st edn). American Psychiatric Press, Washington, DC.
2. American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd edn). American Psychiatric Press, Washington, DC.
3. Marks, I.M. and Gelder, M.G. (1966). Different ages of onset in varieties of social phobia. *The American Journal of Psychiatry*, **123**, 218–21.
4. American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Press, Washington, DC.
5. Liebowitz, M.R., Gorman, J.M., Fyer, A.J., *et al.* (1985). Social phobia: review of a neglected anxiety disorder. *Archives of General Psychiatry*, **42**, 729–36.
6. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Press, Washington, DC.
7. Rapee, R.M. and Heimberg, R.G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, **35**, 741–56.
8. Clark, D.M. and Wells, A. (1995). A cognitive model of social phobia. In *Social phobia: diagnosis, assessment, and treatment* (eds. R.G. Heimberg, M.R. Liebowitz, D.A. Hope, and F.R. Schneier), pp. 69–93. Guilford Press, New York.
9. Schneier, F.R., Heckelman, L.R., Garfinkel, R., *et al.* (1994). Functional impairment in social phobia. *The Journal of Clinical Psychiatry*, **55**, 322–31.
10. Wittchen, H., Fuetsch, M., Sonntag, H., *et al.* (1999). Disability and quality of life in pure and comorbid social phobia: findings from a controlled study. *European Psychiatry*, **14**, 118–31.
11. Whisman, M., Sheldon, C., and Goering, P. (2000). Psychiatric disorders and dissatisfaction with social relationships: does type of relationship matter? *Journal of Abnormal Psychology*, **109**, 803–8.
12. Stein, M., McQuaid, J., Laffaye, C., *et al.* (1999). Social phobia in the primary medical care setting. *Journal of Family Practice*, **49**, 514–9.
13. World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: clinical description and diagnostic guidelines. World Health Organization, Geneva.
14. Heimberg, R.G., Hope, D.A., Dodge, C.S., *et al.* (1990). DSM-III-R subtypes of social phobia: comparison of generalized social phobics and public speaking phobics. *The Journal of Nervous and Mental Disease*, **173**, 172–9.
15. Heimberg, R.G., Holt, C.S., Schneier, F.R., *et al.* (1993). The issue of subtypes in the diagnosis of social phobia. *Journal of Anxiety Disorders*, **7**, 249–69.
16. Vriends, N., Becker, E.S., Meyer, A., *et al.* (2007). Subtypes of social phobia: are they of any use? *Journal of Anxiety Disorders*, **21**, 59–75.
17. Heimberg, R.G. (1996). Social phobia, avoidant personality disorder and the multiaxial conceptualization of interpersonal anxiety. In *Trends in cognitive and behavioural therapies*, Vol. 1 (ed. P. Salkovskis), pp. 43–61. John Wiley & Sons, Chichester, England.
18. Schneier, F.R., Liebowitz, M.R., Beidel, D.C., *et al.* (1998). MacArthur data reanalysis for *DSM-IV*: social phobia. In *DSM-IV sourcebook*, Vol. 4 (eds. T.A. Widiger, A.H. Frances, H.A. Pincus, R. Ross, M.J. First, W. Davis, and M. Kline), pp. 307–28. American Psychiatric Press, Washington, DC.
19. Magee, W.J., Eaton, W.W., Wittchen, H.U., *et al.* (1996). Agoraphobia, simple phobia, and social phobia in the National Comorbidity Survey. *Archives of General Psychiatry*, **53**, 159–68.
20. Kessler, R.C., Chiu, W.T., Demler, O., *et al.* (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**, 617–27.
21. Heckelman, L.R. and Schneier, F.R. (1995). Diagnostic issues. In *Social phobia: diagnosis, assessment, and treatment* (eds. R.G. Heimberg, M.R. Liebowitz, D.A. Hope, and F.R. Schneier), pp. 3–20. Guilford Press, New York.
22. Mannuzza, S., Fyer, A.J., Liebowitz, M.R., *et al.* (1990). Delineating the boundary of social phobia: its relationship to panic disorder and agoraphobia. *Journal of Anxiety Disorders*, **4**, 41–59.
23. Rapee, R.M., Sanderson, W.C., and Barlow, D.H. (1988). Social phobia features across the DSM-III-R anxiety disorders. *Journal of Psychopathology and Behavioral Assessment*, **10**, 287–99.
24. Reich, J., Noyes, R., and Yates, W. (1988). Anxiety symptoms distinguishing social phobia from panic and generalized anxiety disorders. *The Journal of Nervous and Mental Disease*, **176**, 510–3.
25. Kessler, R.C., Berglund, P.D., Demler, O., *et al.* (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**, 593–602.
26. Hazen, A.L. and Stein, M.B. (1995). Clinical phenomenology and comorbidity. In *Social phobia: clinical and research perspectives* (ed. M.B. Stein), pp. 3–41. American Psychiatric Press, Washington, DC.
27. Wang, P.S., Berglund, P., Olfson, M., *et al.* (2005). Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**, 603–13.
28. Olfson, M., Guardino, M., Struening, E., *et al.* (2000). Barriers to treatment of social anxiety. *The American Journal of Psychiatry*, **157**, 521–7.
29. Turk, C.L., Heimberg, R.G., Orsillo, S.M., *et al.* (1998). An investigation of gender differences in social phobia. *Journal of Anxiety Disorders*, **12**, 209–23.
30. Tillfors, M. (2004). Why do some individuals develop social phobia? A review with emphasis on the neurobiological influences. *Nordic Journal of Psychiatry*, **58**, 267–76.
31. Stein, M.B., Chartier, M.J., Hazen, A.L., *et al.* (1998). A direct-interview family study of generalized social phobia. *The American Journal of Psychiatry*, **155**, 90–7.
32. Kendler, K.S., Neale, M.C., Kessler, R.C., *et al.* (1992). The genetic epidemiology of phobias in women: the interrelationship of agoraphobia, social phobia, situational phobia, and simple phobia. *Archives of General Psychiatry*, **49**, 273–81.
33. Kendler, K.S., Karkowski, L.M., and Prescott, C.A. (1999). Fears and phobias: reliability and heritability. *Psychological Medicine*, **29**, 539–53.
34. Kendler, K.S., Myers, J., Prescott, C.A., *et al.* (2001). The genetic epidemiology of irrational fears and phobias in men. *Archives of General Psychiatry*, **58**, 257–65.
35. Stein, M.B., Goldin, P.R., Sareen, J., *et al.* (2002). Increased amygdala activation to angry and contemptuous faces in generalized social phobia. *Archives of General Psychiatry*, **59**, 1027–34.
36. Blanco, C., Schneier, F.R., Schmidt, A., *et al.* (2003). Pharmacological treatment of social anxiety disorder: a meta-analysis. *Depression and Anxiety*, **18**, 29–40.
37. Stein, D.J. and Stahl, S. (2000). Serotonin and anxiety: current models. *International Clinical Psychopharmacology*, **15**(Suppl. 2), S1–6.
38. Schneier, F.R., Liebowitz, M.R., Abi-Dargham, A., *et al.* (2000). Low dopamine D2 receptor binding potential in social phobia. *The American Journal of Psychiatry*, **157**, 457–9.
39. Bruch, M.A., Heimberg, R.G., Berger, P., *et al.* (1989). Social phobia and perceptions of early parental and personal characteristics. *Anxiety Research*, **2**, 57–65.

40. Lieb, R., Wittchen, H.U., Hofler, M., *et al.* (2000). Parental psychopathology, parenting styles, and the risk of social phobia in offspring. *Archives of General Psychiatry*, **57**, 859–66.
41. Erwin, B.A., Heimberg, R.G., Marx, B.P., *et al.* (2006). Traumatic and socially stressful life events among persons with social anxiety disorder. *Journal of Anxiety Disorders*, **20**, 896–914.
42. Reich, J., Goldenberg, I., Vasile, R., *et al.* (1994). A prospective follow-along study of the course of social phobia. *Psychiatry Research*, **54**, 249–58.
43. Reich, J., Goldenberg, I., Goisman, R., *et al.* (1994). A prospective, follow-along study of the course of social phobia: II. testing for basic predictors of course. *The Journal of Nervous and Mental Disease*, **182**, 297–301.
44. Caspi, A., Edler, G.H., and Bem, D.J. (1988). Moving away from the world: life-course patterns of shy children. *Developmental Psychology*, **24**, 824–31.
45. Schwartz, C.E., Snidman, N., and Kagan, J. (1999). Adolescent social anxiety as an outcome of inhibited temperament in childhood. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**, 1008–15.
46. Rosenbaum, J.F., Biederman, J., Hirshfeld, D.R., *et al.* (1991). Behavioral inhibition in childhood: a possible precursor to panic disorder or social phobia. *The Journal of Clinical Psychiatry*, **52**(Suppl.), 5–9.
47. Rodebaugh, T.L., Holaway, R.M., and Heimberg, R.G. (2004). The treatment of social anxiety disorder. *Clinical Psychology Review*, **24**, 883–908.
48. Al-Kubaisy, T., Marks, I.M., Logsdail, S., *et al.* (1992). Role of exposure homework in phobia reduction: a controlled study. *Behavior Therapy*, **23**, 599–621.
49. Turner, S.M., Beidel, D.C., and Jacob, R.G, (1994). Social phobia: a comparison of behavior therapy and atenolol. *Journal of Consulting and Clinical Psychology*, **62**, 350–8.
50. Newman, M.G., Hofmann, S.G., Trabert, W., *et al.* (1994). Does behavioral treatment of social phobia lead to cognitive changes? *Behavior Therapy*, **25**, 503–17.
51. Heimberg, R.G. and Becker, R.E. (2002). Cognitive-behavioral group therapy for social phobia: basic mechanisms and clinical strategies. Guilford Press, New York.
52. Clark, D.M. (2001). A cognitive perspective on social phobia. In *International handbook of social anxiety: concepts, research and interventions relating to the self and shyness* (eds. W.R. Crozier and L.E. Alden LE), pp. 405–30. John Wiley & Sons, Chichester, United Kingdom.
53. Beck, A.T. and Emery, G. (1985). *Anxiety disorders and phobias: a cognitive perspective*. Basic Books, New York.
54. Bouton, M.E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological Psychiatry*, **52**, 976–86.
55. Hope, D.A., Heimberg, R.G., and Bruch, M.A. (1995). Dismantling cognitive-behavioral group therapy for social phobia. *Behaviour Research and Therapy*, **33**, 637–50.
56. Butler, G., Cullington, A., Munby, M., *et al.* (1984). Exposure and anxiety management in the treatment of social phobia. *Journal of Consulting and Clinical Psychology*, **52**, 642–50.
57. Heimberg, R.G., Liebowitz, M.R., Hope, D.A., *et al.* (1998). Cognitive-behavioral group therapy versus phenelzine in social phobia: 12-week outcome. *Archives of General Psychiatry*, **55**, 1133–41.
58. Heimberg, R.G., Dodge, C.S., Hope, D.A., *et al.* (1990). Cognitive behavioral group treatment for social phobia: comparison with a credible placebo control. *Cognitive Therapy and Research*, **14**, 1–23.
59. Taylor, S. (1996). Meta-analysis of cognitive-behavioral treatments for social phobia. *Journal of Behavior Therapy and Experimental Psychiatry*, **27**, 1–9.
60. Gould, R.A., Buckminster, S., Pollack, M.H., *et al.* (1997). Cognitive-behavioral and pharmacological treatment for social phobia: a meta-analysis. *Clinical Psychology: Science and Practice*, **4**, 291–306.
61. Heimberg, R.G., Salzman, D.G., Holt, C.S., *et al.* (1993). Cognitive-behavioral group treatment for social phobia: effectiveness at five-year followup. *Cognitive Therapy and Research*, **17**, 325–39.
62. Liebowitz, M.R., Heimberg, R.G., Schneier, F.R., *et al.* (1999). Cognitive-behavioral group therapy versus phenelzine in social phobia: long-term outcome. *Depression and Anxiety*, **10**, 89–98.
63. Mörtberg, E., Karlsson, A., Fyring, C., *et al.* (2006). Intensive cognitive-behavioral group treatment (CBGT) of social phobia: a randomized controlled study. *Journal of Anxiety Disorders*, **20**, 646–60.
64. Clark, D.M., Ehlers, A., Hackmann, A., *et al.* (2006). Cognitive therapy versus exposure and applied relaxation in social phobia: a randomized controlled trial. *Journal of Consulting and Clinical Psychology*, **74**, 568–78.
65. Clark, D.M., Ehlers, A., McManus, F., *et al.* (2003). Cognitive therapy vs fluoxetine in generalized social phobia: a randomized placebo controlled trial. *Journal of Consulting and Clinical Psychology*, **71**, 1058–67.
66. Stangier, U., Heidenreich, T., Peitz, M., *et al.* (2003). Cognitive therapy for social phobia: individual versus group treatment. *Behaviour Research and Therapy*, **41**, 991–1007.
67. Mörtberg, E., Clark, D.M., Sundin, Ö., *et al.* (2007). Intensive group cognitive treatment and individual cognitive therapy vs treatment as usual in social phobia: a randomized controlled trial. *Acta Psychiatrica Scandinavica*, **115**, 142–54.
68. Fedoroff, I.C. and Taylor, S. (2001). Psychological and pharmacological treatments for social anxiety disorder: a meta-analysis. *Journal of Clinical Psychopharmacology*, **21**, 311–24.
69. Kobak, K.A., Greist, J.H., Jefferson, J.W., *et al.* (2002). Fluoxetine in social phobia: a double-blind, placebo-controlled pilot study. *Journal of Clinical Psychopharmacology*, **22**, 257–62.
70. Lader, M., Stender, K., Burger, V., *et al.* (2004). The efficacy and tolerability of escitalopram in the short-and long-term treatment of social anxiety disorder: a randomised, double blind, placebo-controlled, fixed-dose study. *Depression and Anxiety*, **19**, 241–48.
71. Liebowitz, M.R., Gelenberg, A.J., and Munjack, D, (2005). Venlafaxine extended release vs placebo and paroxetine in social anxiety disorder. *Archives of General Psychiatry*, **62**, 190–8.
72. Liebowitz, M.R., Stein, M.B., Tancer, M., *et al.* (2002). A randomized, double-blind, fixed-dose comparison of paroxetine and placebo in the treatment of generalized social anxiety disorder. *The Journal of Clinical Psychiatry*, **63**, 66–74.
73. Stein, M.B., Liebowitz, M.R., Lydiard, R.B., *et al.* (1998). Paroxetine treatment of generalized social phobia (social anxiety disorder): a randomized controlled trial. *The Journal of the American Medical Association*, **280**, 708–13.
74. Stein, D.J., Versiani, M., Hair, T., *et al.* (2002). Efficacy of paroxetine for relapse prevention in social anxiety disorder: a 24-week study. *Archives of General Psychiatry*, **59**, 1111–18.
75. Van Ameringen, M.A., Lane, K.M., Walker, J.R., *et al.* (2001). Sertraline treatment for generalized social phobia: a 20-week, double-blind, placebo controlled study. *The American Journal of Psychiatry*, **158**, 275–81.
76. Walker, J.R., van Ameringen, M.A., Swinson, R., *et al.* (2000). Prevention of relapse in generalized social phobia: results of a 24-week study in responders to 20 weeks of sertraline treatment. *Journal of Clinical Psychopharmacology*, **20**, 636–44.
77. Stein, M.B., Fyer, A.J., Davidson, J.R.T., *et al.* (1999). Fluvoxamine treatment of social phobia (social anxiety disorder): a double-blind, placebo-controlled study. *The American Journal of Psychiatry*, **156**, 756–60.
78. Davidson, J.R., Foa, E.B., Huppert, J.D., *et al.* (2004). Fluoxetine, comprehensive cognitive behavioral therapy, and placebo in generalized social phobia. *Archives of General Psychiatry*, **61**, 1005–13.
79. Davidson, J.R.T., Potts, N., Richichi, E., *et al.* (1993). Treatment of social phobia with clonazepam and placebo. *Journal of Clinical Psychopharmacology*, **13**, 423–8.

80. Otto, M.W., Pollack, M.H., Gould, R.A., *et al.* (2000). A comparison of the efficacy of clonazepam and cognitive-behavioral group therapy for the treatment of social phobia. *Journal of Anxiety Disorders*, **14**, 345–58.
81. Gelernter, C.S., Uhde, T.W., Cimbolic, P., *et al.* (1991). Cognitive-behavioral and pharmacological treatments of social phobia: a controlled study. *Archives of General Psychiatry*, **48**, 938–45.
82. Liebowitz, M.R., Schneier, F.R., Campeas, R., *et al.* (1992). Phenelzine vs atenolol in social phobia: a placebo-controlled comparison. *Archives of General Psychiatry*, **49**, 290–300.
83. Versiani, M., Nardi, A.E., Mundim, F.D., *et al.* (1992). Pharmacotherapy of social phobia: a controlled study with moclobemide and phenelzine. *The British Journal of Psychiatry*, **161**, 353–60.
84. Clark, D.B. and Agras, W.S. (1991). The assessment and treatment of performance anxiety in musicians. *The American Journal of Psychiatry*, **148**, 598–605.
85. Schneier, F.R., Saoud, J., Campeas, R., *et al.* (1993). Buspirone in social phobia. *Journal of Clinical Psychopharmacology*, **13**, 251–6.
86. Pande, A.C., Davidson, J.R.T., Jefferson, J.W., *et al.* (1999). Treatment of social phobia with gabapentin: a placebo-controlled study. *Journal of Clinical Psychopharmacology*, **19**, 341–8.
87. Feltner, D.E., Pollack, M.H., Davidson, J.R.T., *et al.* (2000). A placebo-controlled, double-blind study of pregabalin treatment of social phobia: outcome and predictors of response. *European Neuropsychopharmacology*, **10**, 345–55.
88. Walker, D.L., Ressler, K.J., Lu, K.T., *et al.* (2002). Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of d-cycloserine as assessed with fear-potentiated startle in rats. *The Journal of Neuroscience*, **22**, 2343–51.
89. Hofmann, S.G., Meuret, A.E., Smits, J.A.J., *et al.* (2006). Augmentation of exposure therapy with d-cycloserine for social anxiety disorder. *Archives of General Psychiatry*, **63**, 298–304.
90. Blomhoff, S., Haug, T.T., Hellstrøm, K., *et al.* (2001). Randomized controlled general practice trial of setraline, exposure therapy, and combined treatment in generalized social phobia. *The British Journal of Psychiatry*, **179**, 23–30.
91. Haug, T.T., Blomhoff, S., Hellstrøm, K., *et al.* (2003). Exposure therapy and sertraline in social phobia: 1-year follow-up of a randomized controlled trial. *The British Journal of Psychiatry*, **182**, 312–8.
92. Heimberg, R.G. (2002). Cognitive behavioral therapy for social anxiety disorder: current status and future directions. *Biological Psychiatry*, **51**, 101–8.
93. Safren, S.A., Heimberg, R.G., and Juster, H.R. (1997). Clients' expectancies and their relationship to pretreatment symptomatology and outcome of cognitive-behavioral group treatment for social phobia. *Journal of Consulting and Clinical Psychology*, **65**, 694–8.
94. Erwin, B.A., Heimberg, R.G., Juster, H., *et al.* (2002). Comorbid anxiety and mood disorders among persons with social anxiety disorder. *Behaviour Research and Therapy*, **40**, 19–35.
95. Barrett, P.M., Rapee, R.M., Dadds, M.R., *et al.* (1996). Family enhancement of cognitive style in anxious and aggressive children. *Journal of Abnormal Child Psychology*, **24**, 187–203.
96. Barrett, P. and Turner, C. (2001). Prevention of anxiety symptoms in primary school children: preliminary results from a universal school-based trial. *The British Journal of Clinical Psychology*, **40**, 399–410.
97. Dadds, M.R., Holland, D.E., Laurens, K.R., *et al.* (1999). Early intervention and prevention of anxiety disorders in children: results at 2-year follow-up. *Journal of Consulting and Clinical Psychology*, **67**,145–50.
98. Barlow, D.H., DiNardo, P.A., Vermilyea, B.B., *et al.* (1986). Co-morbidity and depression among the anxiety disorders: issues in diagnosis and classification. *The Journal of Nervous and Mental Disease*, **174**, 63–72.
99. Merckelbach, H., de Jong, P.J., Muris, P., *et al.* (1996). The etiology of specific phobias: a review. *Clinical Psychology Review*, **16**, 337–61.
100. Thyer, B.A., Himle, J., and Curtis, G.C. (1985). Blood-injury-illness phobia: a review. *Journal of Clinical Psychology*, **41**, 451–9.
101. Fredrikson, M., Annas, P., Fischer, H., *et al.* (1996). Gender and age differences in the prevalence of specific fears and phobias. *Behaviour Research and Therapy*, **34**, 33–9.
102. Scheibe, G. and Albus, M. (1992). Age at onset, precipitating events, sex distribution, and co-occurrence of anxiety disorders. *Psychopathology*, **25**, 11–8.
103. Öst, L.G. and Treffers, P.D.A. (2001). Onset, course, and outcome for anxiety disorders in children. In *Anxiety disorders in children and adolescents: research, assessment and intervention* (eds. W.K. Silverman and P.D.A. Treffers), pp. 293–312. Cambridge University Press, Cambridge.
104. Fyer, A.J., Mannuzza, S., Chapman, T.F., *et al.* (1995). Specificity in familial aggregation of phobic disorders. *Archives of General Psychiatry*, **52**, 564–73.
105. Fyer, A.J., Mannuzza, S., Gallops, M.S., *et al.* (1990). Familial transmission of simple phobias and fears: a preliminary report. *Archives of General Psychiatry*, **47**, 252–6.
106. Biederman, J., Rosenbaum, J.F., Hirshfeld, D.R., *et al.* (1990). Psychiatric correlates of behavioral inhibition in young children of parents with and without psychiatric disorders. *Archives of General Psychiatry*, **47**, 21–6.
107. Pavlov, I. (1927). *Conditioned reflexes*. Oxford University Press, London.
108. Mowrer, O.H. (1939). Stimulus response theory of anxiety. *Psychological Review*, **46**, 553–65.
109. Marks, I.M. (1969). *Fears and phobias*. Academic Press, New York.
110. Seligman, M.E. (1971). Phobias and preparedness. *Behavior Therapy*, **2**, 307–20.
111. Boyd, J.H., Rae, D.S., Thompson, J.W., *et al.* (1990). Phobia: prevalence and risk factors. *Social Psychiatry and Psychiatric Epidemiology*, **25**, 314–23.
112. Marks, I.M. (1987). Fears, phobias, and rituals: panic, anxiety, and their disorders. Oxford University Press, London.
113. Hecker, J.E. (1990). Emotional processing in the treatment of simple phobia: a comparison of imaginal and in vivo exposure. *Behavioural Psychotherapy*, **18**, 21–34.
114. Gotestam, K.G. and Berntzen, D. (1997). Use of the modeling effect in one-session exposure. *Scandinavian Journal of Behaviour Therapy*, **26**, 97–101.
115. Öst, L.G. (1996). One-session group treatment of spider phobia. *Behaviour Research and Therapy*, **34**, 707–15.
116. Chambless, D.L. (1990). Spacing of exposure sessions in treatment of agoraphobia and simple phobia. *Behavior Therapy*, **21**, 217–29.
117. Öst, L.G., Salkovskis, P.M., and Hellstrom, K. (1991). One-session therapist-directed exposure vs self-exposure in the treatment of spider phobia. *Behavior Therapy*, **22**, 407–22.
118. Carlin, A.S., Hoffman, H.G., and Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. *Behaviour Research and Therapy*, **35**, 153–8.
119. North, M.M., North, S.M., and Coble, J.R. (1997). Virtual reality therapy for fear of flying. *The American Journal of Psychiatry*, **154**, 130.
120. Rothbaum, B.O., Hodges, L., Watson, B.A., *et al.* (1996). Virtual reality exposure therapy in the treatment of fear of flying: a case report. *Behaviour Research and Therapy*, **34**, 477–81.
121. Rothbaum, B.O., Hodges, L., Anderson, P.L., *et al.* (2002). Twelve-month follow-up of virtual reality and standard exposure therapies for the fear of flying. *Journal of Consulting and Clinical Psychology*, **70**, 428–32.
122. Maltby, N., Kirsch, I., Mayers, M., *et al.* (2002). Virtual reality exposure therapy for the treatment of fear of flying: a controlled investigation. *Journal of Consulting and Clinical Psychology*, **70**, 1112–8.
123. Ressler, K.J., Rothbaum, B.O., Tannenbaum, L., *et al.* (2004). Cognitive enhancers as adjuncts to psychotherapy: use of d-cycloserine in phobic individuals to facilitate extinction of fear. *Archives of General Psychiatry*, **61**, 1136–44.

124. Bernstein, D.A., Borkovec, T.D., and Hazlett-Stevens, H. (2000). *New directions in progressive relaxation training: a guidebook for helping professionals*. Prager/Greenwood, Westport, CT.
125. Öst, L.G. (1987). Applied relaxation: description of a coping technique and review of controlled studies. *Behaviour Research and Therapy*, **25**, 397–409.
126. Öst, L.G., Johansson, J., and Jerremalm, A. (1982). Individual response patterns and the effects of different behavioral methods in the treatment of claustrophobia. *Behaviour Research and Therapy*, **20**, 445–60.
127. Öst, L.G., Sterner, U., and Fellenius, J. (1989). Applied tension, applied relaxation, and the combination in the treatment of blood phobia. *Behaviour Research and Therapy*, **27**, 109–21.
128. Öst, L.G., Fellenius, J., and Sterner, U. (1991). Applied tension, exposure *in vivo*, and tension-only in the treatment of blood phobia. *Behaviour Research and Therapy*, **29**, 561–74.
129. Hellstrøm, K., Fellenius, J., and Öst. L.G. (1996). One versus five sessions of applied tension in the treatment of blood phobia. *Behaviour Research and Therapy*, **34**, 101–12.
130. Thorpe, S.J. and Salkovskis, P.M. (1995). Phobia beliefs: do cognitive factors play a role in specific phobias? *Behavior Research and Therapy*, **33**, 805–16.
131. Greco, T.S. (1989). A cognitive-behavioral approach to fear of flying: a practitioner's guide. *Phobia Practice and Research Journal*, **2**, 3–15.
132. Campos, P.E., Solyom, L., and Koelink, A. (1984). The effects of timolol maleate on subjective and physiological components of air travel phobia. *Canadian Journal of Psychiatry*, **29**, 570–4.
133. Sartory, G. (1983). Benzodiazepines and behavioral treatment of phobic anxiety. *Behavioural Psychotherapy*, **11**, 204–17.
134. Norton, G.R. and Johnson, W.E. (1983). A comparison of two relaxation procedures for reducing cognitive and somatic anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, **14**, 209–14.

4.7.3 Panic disorder and agoraphobia

James C. Ballenger

Introduction

Panic disorder draws its name from the Greek god Pan, god of flocks. Pan was known for suddenly frightening animals and humans 'out of the blue'. The spontaneous 'out of the blue' character of panic attacks is the principal identifying characteristic of panic disorder and central to its recognition and diagnosis.

We know the syndrome that we currently call panic disorder with and without agoraphobia has probably existed since the beginning of recorded history. Hippocrates presented cases of obvious phobic avoidance around 400 BC.[1] One of the first modern descriptions was by Benedikt around 1870, describing individuals who developed sudden anxiety and dizziness in public places.[2]

Certainly, our current modern ideas of panic disorder evolved essentially simultaneously in the United States and Europe in the early to mid-1960s. Donald Klein in the United States described in 1964 the panic syndrome and reported that it was responsive to imipramine.[3] Isaac Marks in the United Kingdom also described panic attacks and agoraphobic avoidance, and treating the syndrome effectively with behaviour therapy.[4]

Until the last several decades, panic disorder and agoraphobia were actually thought to be rare syndromes. It is now clear that individuals with these difficulties are anything but rare. In fact, panic disorder is one of the most common presenting problems in individuals seeking mental health treatment and the fifth most common problem seen in primary care settings.[5] It was thought to be a mild problem, but we now know that it is associated with significant dysfunction. The disability in social, occupational, and family life is in fact comparable to major depression.

Although there are differences in the understandings of panic disorder and its treatments across the world, this chapter will review the current understanding about panic disorder, its characteristics, diagnosis, aetiology, and treatments.

Clinical features

One of the earliest and most accurate descriptions of panic attacks was provided by Charles Darwin in 1872 as he described his own episodes:

> The heartbeats quickly and violently so that it palpitates and knocks against the ribs . . . the skin instantly becomes pale as during incipient faintness . . . under a sense of great fear . . . in connection with the disturbed action of the heart, the breathing is hurried . . . one of the best marked symptoms is the trembling of all the muscles of the body.[6]

The most characteristic type of panic attack is the spontaneous 'out of the blue' episode of extreme anxiety. Other 'situational panic attacks' occur immediately upon exposure, or in anticipation of exposure to particular situations, usually where panic attacks have occurred previously. Some individuals have panic attacks in certain situations some of the time, but not always, and these are labelled 'situationally predisposed panic attacks'.

Panic attacks also occur in other anxiety syndromes and are more or less the same in whatever syndrome where they occur. However, spontaneous panic attacks in panic disorder tend to have more dizziness, paraesthesia, shaking, chest pain, and fears of going mad. Shortness of breath is more common in panic attacks in agoraphobia. Certainly blushing is particularly characteristic of panic attacks in social phobia.

The symptoms of panic attacks in order of their frequency include palpitations, pounding heart, tachycardia, sweating, trembling or shaking, shortness of breath or smothering, feeling of choking, chest pain or discomfort, nausea or abdominal distress, feeling dizzy, unsteady, lightheaded or faint, derealization or depersonalization, fear of losing control or going mad, fear of dying, paraesthesia, and chills or hot flushes.

Panic attacks by definition generally involve four or more of the above symptoms to meet diagnostic criteria for panic disorder in the DSM-IV. The anxiety is crescendo in nature, building to a peak in 10 min in most cases. Panic attacks usually last for several minutes, but in some patients they can last for hours. The frequency and severity of panic attacks varies greatly between individuals and, at times, in individuals. Some have only one to three panic attacks per year, whereas others may have multiple panic attacks each day. Some individuals have bursts of panic attacks and then an absence of all attacks for a period of time. Across a large panic disorder clinical trial, the typical patient described one to two panic attacks per week.[7]

Panic attacks with fewer than four symptoms have been labelled 'limited-symptom attacks' or 'little panic attacks', and most individuals with panic disorder have these, as well as panic attacks with four or more symptoms. The threshold of four symptoms was chosen for DSM-IV because individuals with panic attacks with four symptoms or more had more disability than patients with one- to three-symptom attacks. This threshold is clearly arbitrary, and patients having panic attacks with fewer symptoms do have significant morbidity.

Panic attacks are extremely frightening and patients develop an essentially logical fear of having more panic attacks. Patients develop worry and anxiety about the possibility of panic attacks recurring. This anxiety between attacks has been called 'anticipatory anxiety' and can be almost constant, and characteristically increases prior to exposure to situations previously associated with panic attacks (e.g. having to shop in the supermarket where a panic attack has occurred).

A significant number of people with panic attacks go on to develop fear and avoidance of situations associated with previous panic attacks. They also fear situations where escape would be difficult or embarrassing, or where help might not be available. Most patients mistakenly believe they become incapacitated and incapable of taking care of themselves during a panic attack and therefore, many go on to develop avoidance of a variety of situations where they could not easily get help. Factor analytical studies demonstrate that there are clusters of situations associated with avoidance. These typically include public transportation (e.g. buses, trains, planes), riding in or driving a car, especially on heavily travelled roads, crowds (e.g. the cinema, a football match, large shopping centres), shopping (especially in supermarkets), particularly where one must stand in queues, and bridges, tunnels, elevators, and other enclosed spaces.[8] In the event of a panic attack, people often have an overwhelming need to escape or return to a place of safety such as home. Therefore they fear situations where escape is difficult or impossible, e.g. airplanes, traffic jams on a bridge, dental appointments, etc. On closer examination, it is clear that patients do not actually fear the situation itself but rather reason 'what if' the panic feelings occur while in that situation. This has led to the syndrome being called the 'what if' syndrome, emphasizing that there is actually a 'fear of the fear'.[9]

Patients tend to avoid such situations or force themselves to endure them in distress or take a companion along 'to help'. Others limit their travel to short distances from home or take longer routes where they perceive help would be available (e.g. police, doctors' offices, fire stations, etc).

Some patients develop agoraphobic avoidance following their first attack, some only after frequent and severe attacks, and some never develop agoraphobic avoidance. In community samples, one-third to half of patients who meet criteria for panic disorder also has significant agoraphobic avoidance. The rate is higher (75 per cent) in most clinical samples. A minority (5 per cent) ultimately become unable to leave their homes and are housebound.

Many patients have panic attacks that awaken them from sleep (nocturnal panic attacks), as well as during the day. These are in fact quite common and the majority of panic disorder patients experience them. These occur during slow wave sleep early in sleep. These panic attacks are essentially identical to traditional panic attacks that occur during the day. There is a group of patients who have what are called non-fearful panic attacks. These involve the sudden onset of physiological symptoms without the cognitive components of fear or anxiety. These primarily are medical patients, usually cardiac, who might have episodes of sudden tachycardia and palpitations but no fear.[10]

Classification

The earliest modern accounts of what is almost certainly the panic disorder syndrome began appearing in the mid-1900s. There were accounts beginning in 1866 of paroxysmal anxiety episodes that did not use the term 'panic attack'. During the American Civil War, patients were diagnosed with 'irritable heart syndrome' or 'Da Costa syndrome' (1871) with clear descriptions of what we now know as panic disorder. Westphal in Germany in 1872 clearly described patients having panic attacks and agoraphobic avoidance of wide open spaces.[11] Again, in the First World War (1918) the syndrome 'neurocirculatory asthenia' was described which had most of the features of panic disorder.

It was, in fact, Freud in Case IV of Katharina, published in 1895, who set the background for the modern classification of panic disorder. However, it was the Feighner criteria published in the United States in 1972 that give the first formal diagnostic recognition to the syndrome.[12] The Research Diagnostic Criteria (RDC) which followed in 1978 first split panic disorder away from what we now call generalized anxiety disorder (GAD). In the RDC, panic disorder had panic attacks while GAD did not. These diagnoses were made part of the DSM-III diagnostic scheme in 1980.

It was the conceptualization of panic disorder by Donald Klein in 1964 in the United States that led to the predominant view of the syndrome, certainly in the United States.[13] Klein argued that panic attacks were the core of the syndrome, and the remaining clinical phenomena were consequences of the panic attacks. He conceptualized that anticipatory anxiety was the fear of the possible recurrence of panic attacks, and similarly that agoraphobia was the subsequent fear and avoidance of situations where panic attacks had occurred. The bringing together of these three phenomena into one concept was accepted in the DSM-IIIR, and more recently in the DSM-IV in the United States.[14]

The biological findings that typical panic attacks could be elicited in panic disorder patients with infusions of lactate, doses of caffeine, or breathing 35 per cent CO_2 supported this conceptualization of the syndrome as primarily centered around panic attacks. This hypothesis was further supported by epidemiological findings of essentially the same illness around the world.

However, this idea remains controversial across the different sides of the Atlantic. The American DSM-IIIR and DSM-IV diagnostic schema continue to utilize the idea that panic attacks are pre-eminent and created two diagnoses: panic disorder and panic disorder with agoraphobia. In Europe and in the ICD-10, agoraphobia is conceptualized as dominant over panic attacks. Therefore, when agoraphobia and panic attacks are both present, the diagnosis is conceptualized as a phobia and that condition is diagnosed as agoraphobia with panic attacks. Beyond this theoretical debate is the clinical question whether the treatment should be aimed first at panic attacks (in the United States concept) or at agoraphobic avoidance, for example with exposure therapy (in the European schema).

(a) Diagnosis

The most recently revised diagnostic schema for this syndrome is the DSM-IV. Changes from the DSM-IIIR were made based on two principles:

1 any new empirical data that required changes be made;

2 an attempt to make the DSM-IV and ICD-10 more compatible.

For the diagnoses of panic disorder and agoraphobia, an attempt was also made to more nearly describe the prototypic patient and to move away from the pseudoquantification of using number of panic attacks per week.[14]

The DSM-IV clarified that panic attacks occurred in multiple syndromes including social phobia, obsessive–compulsive disorder, depression, and others. However, DSM-IV utilized the distinction that only in panic disorder were there recurrent spontaneous panic attacks not bound to any particular situation. The diagnosis of panic disorder has several requirements including the following:

- Recurrent, unexpected panic attacks (situational panic attacks could also occur but there would need to be at least two unexpected panic attacks by history).
- Panic attacks needed to be followed by at least 1 month of persistent anxiety about potential recurrence of further panic attacks, implications of these attacks (e.g. going mad, something wrong medically), or a significant behavioural change because of these attacks. This was necessary because some patients had panic attacks and completely changed their lives but denied that they were worried about experiencing more panic attacks or the implications of the panic attacks.[14]

The agoraphobia criteria remained largely unchanged, but it was made more clear that the diagnosis was based on persistent fear and avoidance of certain clusters of situations and listed the most common.

The controversial diagnosis of agoraphobia without a history of panic disorder was retained until further clarification is obtained through research. Our current understanding is that these patients generally have never fully met criteria for panic disorder because their panic-like symptoms have not met the diagnostic criterion requiring four full symptoms for a panic attack. Available research suggests that these patients are otherwise very much like typical patients with panic disorder and agoraphobia. Some patients actually have only one or two symptoms (e.g. fear of loss of bladder or bowel control or only tachycardia).

Perhaps the most difficult diagnostic issue is the frequent comorbidity. The Epidemiologic Catchment Area study documented that approximately 50 per cent of panic disorder patients over their lifetime have another anxiety disorder.[15]

In actuality, depression is more commonly comorbid with panic disorder than even agoraphobia and suggests a close relationship between these syndromes. Comorbid depression ranged from 22.5 to 68.2 per cent in various samples. Lifetime rates vary from 35 to 91 per cent.[16] Although approximately half the patients developed panic disorder and depression at essentially the same time, one-quarter develop depression before panic disorder and one-quarter panic disorder before depression.[17] Surprisingly, bipolar illness has been reported to be as high as 20.8 per cent.

Easily one-third of panic disorder patients abuse alcohol. The percentage in clinical samples is much higher with 13 to 43 per cent of panic disorder patients also meeting criteria for alcoholism.[18]

(b) Differential diagnosis

It is particularly important to determine whether agoraphobic avoidance is present, because its treatment usually requires some sort of exposure therapy. Patients will often not volunteer that they are avoiding certain situations out of embarrassment. As mentioned earlier, depression and panic disorder often occur together and again patients often do not describe the other syndrome, but rather describe the syndrome which is most painful to them at that time. However, proper recognition of comorbid depression is especially important because of the marked increase (four-fold) in suicide attempts in patients with panic disorder and depression.

The difference between panic disorder and GAD depends on whether patients have panic attacks and whether they have multiple, unrealistic, and excessive worries about most aspects of life, not just panic attacks. These worries in GAD often concern money, health, children, work problems, etc. The differential with social phobia centres on whether the anxiety is confined entirely towards social situations where the individual fears embarrassment and humiliation. Specific phobias involve panic attacks, but they occur in very specific situations (e.g. high places, thunderstorms) or in the presence of specific objects (e.g. animals, snakes). The post-traumatic stress disorder patient may have many panic-like symptoms, but their illness begins quite specifically after a traumatic experience and anxiety is associated with reminders of that trauma. Finally, obsessive–compulsive disorder can involve panic attacks but only in the specific context of obsessional concerns (e.g. contamination, etc.). In these patients, panic attacks are also dwarfed in importance by typical obsessions and compulsions/rituals concerning contamination, symmetry, bad events, etc.

(c) Medical conditions

Panic-like symptoms do occur in various medical conditions (hyperthyroidism, phaeochromocytoma, hyperparathyroidism, seizures, cardiac arrhythmias, especially supraventricular tachycardia, inner ear difficulties, chronic obstructive pulmonary disease, use of marijuana, withdrawal from drugs of abuse, and over the counter drugs containing caffeine or pseudoephedrine) (Table 4.7.3.1). Also, the typical panic disorder patient does report a large number of physical symptoms and usually to a non-psychiatric physician.

As mentioned, there are medical conditions that can mimic panic disorder (Table 4.7.3.1). There is also evidence that there are slightly increased rates of certain illnesses, for example hyperthyroidism (11–13 per cent) and perhaps mitral valve prolapse. However, these are uncommon in panic disorder patients. Most experts recommend a relatively conservative diagnostic medical work-up. Generally, the most valuable part of a medical examination is a careful history with follow-up of any strongly suggested possibilities, supplemented by a few laboratory tests (complete blood count, thyroid function tests, metabolic screen, and ECG, especially if the patient is over 40 years of age).

(d) Panic disorder in the general medical setting

Conservative estimates of panic disorder in primary care have ranged from 3 to 8 per cent with at least 50 per cent going unrecognized.[19] The average panic disorder patient in general medicine

Table 4.7.3.1 Medical conditions that produce panic-like symptoms

Endocrine	*Respiratory*
Hyperthyroidism	Chronic obstructive
Hypoparathyroidism	pulmonary disease
Hypoglycaemia	Asthma
Phaeochromacytoma	*Substance-induced*
Carcinoid syndrome	Caffeine
Cushing's disease	Cocaine
Cardiovascular	Marijuana
Arrhythmias (supraventricular)	Theophylline
Atypical chest pain	Amphetamines
Mitral valve prolapse	Steroids
Neurological	Alcohol/sedative
Seizures	withdrawal
Vestibular disease	*Haematological*
	Anaemia

takes 10 years or more for a correct diagnosis to be made with an escalation of the use of health care services over this period. In general, the presence of panic disorder leads to a three-fold increase in utilization of general medical services.

The percentage of panic disorder patients is also markedly higher in certain medical groups. These include the very prevalent but most difficult to diagnose patients with vague symptoms such as fatigue, back pain, headache, dizziness, chest pain, etc.[20] or multiple symptoms (more than five).[21]

In a classic study of unrecognized panic disorder patients who were referred for a psychiatric consultation from primary care, 39 per cent had cardiovascular symptoms, 33 per cent gastrointestinal symptoms, and 44 per cent neurological.[22] It is now clear that 16 to 25 per cent of patients presenting to the emergency room with chest pain have panic disorder. Fully 25 per cent of cardiology practice involves panic disorder, usually unrecognized with 80 per cent of patients with chest pain and normal angiograms ultimately diagnosed with panic disorder. We could also increase our yield of recognizing panic disorder patients in certain procedure-oriented situations. For instance, 28 per cent of patients referred for Holter monitoring for palpitations have panic disorder, as do 66 per cent of patients undergoing a work-up to rule out phaeochromocytoma.[23] Also, 44 per cent of irritable bowel syndrome and 15 per cent of patients with headache symptoms seeing a physician have panic disorder, and these are both very prevalent disorders.

Recent studies document that treatment of panic disorder in the medical setting when diagnosed there is most cost-effective.

(e) Comment

A recent large study sponsored by the World Health Organization (WHO) studied primary care patients in 14 different countries.[19] Of patients in that study who had a single panic attack in the previous month, 99 per cent had an anxiety disorder or depression, or a subthreshold anxiety disorder or depressive disorder. The occurrence of a single panic attack also predicted the onset of panic disorder in two-thirds of the patients studied in the next year, a four-fold increase in depression (51 per cent) in the next year, and marked increases in alcoholism, social phobia, and obsessive–compulsive disorder. It would appear that the occurrence of even a single panic attack may well represent the 'tip of the iceberg' and should serve as a signal for increased scrutiny for anxiety and depressive syndromes. This has been recently replicated in a large ($N = 3021$) European sample.[24]

Epidemiology

Surveys largely utilizing DSM diagnoses have found wide agreement and generally equal prevalence's of panic disorder across many countries.[16,25] Utilizing specific criteria for panic attacks, prevalence for panic attacks has generally averaged 7 to 9 per cent of the population (range 1.8–22.7 per cent). However, if criteria for panic attacks are liberalized somewhat ('fearful spells') in terms of the number of times and severity, the prevalence doubles.

There is a striking uniformity worldwide for the observed prevalence of panic disorder. In 10 community studies involving over 40 000 subjects, the majority of studies found lifetime prevalence rates for panic disorder averaging 1.5 to 3.7 per cent, with a yearly prevalence of around 1 and 1.1 per cent of panic disorder with agoraphobia (lifetime).[25] In clinical samples there is greater variability. In the previously mentioned WHO survey of 14 countries, the prevalence for panic disorder in primary care ranged from a low of 1.4 per cent to a high of 16.5 per cent for panic attacks and 0 to 3.5 per cent for panic disorder itself.[19] The average was 1.1 per cent (currently) and 3.5 per cent (lifetime), which is surprisingly similar to the community samples. As mentioned, rates are much higher in specialized medical clinics and range from 15 per cent in dizziness clinics, to 16 to 65 per cent in cardiology practices, to 35 per cent in hyperventilation clinics, etc.

Risk factors

Panic disorder has been uniformly observed to be at least two times more prevalent in females than males.[25] The Epidemiologic Catchment Area study demonstrated a prevalence of 3:2. In clinical samples it is generally 3:1. The onset of panic disorder appears to fall into two peaks. The first occurs in the early to mid-twenties (15–24 years old) with a second peak at 45–54 years of age. The onset of panic disorder after the age of 65 is rare (0.1 per cent).

The highest rates of panic disorder and agoraphobia occur in widowed, divorced, or separated individuals living in cities. Limited education, early parental loss, and physical or sexual abuse are also risk factors. Agoraphobia is clearly more prevalent in females, and females make up three-quarters of the sample with extensive avoidance. Males tend to have longer duration of illness but less agoraphobia and depression, and less frequent help seeking. Perhaps the greater necessity to perform in the workplace retards avoidance in males.

Aetiology

Genetic predisposition

Certainly the preponderance of evidence suggests there is a genetic contribution to the predisposition to develop panic attacks and agoraphobia. There are increased rates of panic disorder in first-degree relatives ranging from 2- to 20-fold with the median

seven- to eight-fold. Overall, studies suggest that another affected relative can be found 25 to 50 per cent of the time, two times as often in female relatives. The increased familial aggregation is specific for panic disorder. These findings are certainly consistent with a modest genetic transmission with relatively high specificity.

Although twin studies are limited, Torgersen[26] did find increased concordance in monozygotic twins (31 versus 0 per cent). In the largest sample of interviewed female twins, a several-fold increase was again found (23.9 versus 10.9 per cent).[27] Skre *et al.*[28] found a two-fold increase of panic disorder in monozygotic twins, while other studies fail to find an increased incidence.

Overall, evidence from family and twin studies suggests that panic disorder involves modest inheritability of around 30 to 40 per cent. The best model suggests 50 per cent genetic and 50 per cent environmental influences. Recent linkage studies to confirm these hypotheses have been contradictory (e.g. with angiotensin, brain-derived neurotropic factor) but do suggest that single-gene transmission is unlikely. However, research is active in this area with positive replicated linkages with chromosomes 13q, 22q, 7p, and 9q31. Identified candidate genes include the ADOR2A, 10832/T, CCK genes, and genes coding for the 5HT1A, 5HT2A, and COMT genes and there is evidence of a link to the corticotrophin releasing hormone gene.[29] This leaves the possibilities of either heterogeneity across families and/or a polygenic inheritance.[30]

Several converging lines do link childhood anxiety with adult anxiety, consistent with a genetic predisposition. This is particularly true for separation anxiety in children. Kagan *et al.*[31] have prescribed that 10 per cent of Caucasian children are born with heightened anxiety which they call behavioural inhibition. Behavioural inhibition is higher in children of anxiety-disordered parents, and there are high rates of anxiety disorders in children of adults with panic disorder. As behaviourally inhibited children have matured, they have been found to show higher rates of anxiety and phobic disorders.[32] Currently, this is probably the best model of an inherited anxiety predisposition. A variant of this type model proposes that there is an aethological factor involving an evolutionarily determined vulnerability to unfamiliar territory. This might explain why the seemingly inherited anxiety is to specific situations. This is also consistent with the high rate of precipitating events prior to the onset of clinical difficulties. In this model an evolutionarily/genetically determined vulnerability would be clinically 'activated' by critical stressors.

Precipitating events have been reported in 60 to 96 per cent of cases. These have often centred on separation or loss, relationship difficulties, taking on new responsibility, and physiological stressors (e.g. childbirth, surgery, hyperthyroidism).[33] This is certainly consistent with a diathesis/vulnerability model with the illness being precipitated in a predisposed individual in adulthood.

There are also many studies suggesting that traumatic early events may figure in the vulnerability leading to panic disorder. The majority of children in some studies have experienced early parental separation. A traumatic event in childhood has been retrospectively reported in at least two-thirds of individuals, a three-fold increase.[34] The most common adult disorder following sexual abuse before the age of five is in fact a 44 per cent incidence of agoraphobia.[35]

There is some evidence in a prospective study involving over 3000 individuals that dependent personality traits were later associated with the development of anxiety disorders. There are also retrospective data that adult panic disorder patients describe their parents as being overly protective and less caring. It is difficult to separate the effects on individuals of the anxiety disorders themselves which create dependent behaviour, or overprotectiveness in the parents producing dependent personality traits.

Biological models

(a) Noradrenaline

There is considerable evidence implicating the brain noradrenaline (norepinephrine) brain systems and panic disorder. The noradrenergic agents yohimbine and isoproterenol stimulate panic attacks in panic disorder patients, suggesting a possible subsensitivity of pre-synaptic alpha 2 inhibitory adrenoreceptors. Both these drugs increase the firing rate of the locus ceruleus, generally thought to be part of the brain anxiety circuit. It is also true that most effective medications in the treatment of panic disorder in fact decrease locus ceruleus firing rate and most panicogenic stimuli increase the locus ceruleus firing rate.

(b) Serotonin (5HT)

Findings with 5HT brain systems in panic disorder are contradictory, probably because of the different 5HT circuits and receptors in different areas of the brain. Most investigators believe, however, that an increase in 5HT transmission decreases panic disorder perhaps because 5HT neurones in ventrolateral periaqueductal grey appear to inhibit sympathoexcitation and the fight or flight response in the rat.[36] The principal human evidence for 5HT being central in panic is that the selective serotonin reuptake inhibitors are effective and that they increase 5HT transmission after long-term use. Also, rapid depletion of 5HT has been shown to result in an increase in panic responses to flumazenil. Whether this increased neurotramsmission in fact leads to downregulation of a supersensitive post-synaptic receptor is one logical possibility, but is as yet unproven. These theories received recent support from PET scan studies demonstrating reductions in brain 5HT1A receptors[37] and 5HT transporter binding.[38]

(c) γ-Aminobutyric acid

The γ-aminobutyric acid (GABA) system is almost certainly involved in panic disorder with perhaps the strongest evidence being that benzodiazepine agonists such as alprazolam and clonazepam are clearly effective treatments for panic disorder. Also GABA antagonists such as flumezanil have panicogenic effects in panic disorder patients, and reverse benzodiazepine agonists such as β-carbolines can cause panic attacks. There is an impaired GABA neuronal response to benzodiazepines (BZs) on brain magnetic spectroscopy and decreased GABA levels in the cingulate and basal ganglia, also on magnetic spectroscopy. Also, positron emission tomography data have demonstrated decreased benzodiazepine binding in the inferior brain areas, including the inferior parieto-temporo-occipital areas.

(d) Cholecystokinin–pentagastrin

Cholecystokinin is clearly involved in anxiety in animals. Also, panic disorder patients develop panic attacks in a dose-dependent fashion with administration of pentagastrin. However, cholecystokinin antagonists have not yet been shown to be effective in humans.

Recent genetic studies implicate CCK gene polymorphisms in panic disorder.

(e) Brain imaging

The explosion of brain imaging data demonstrating a brain circuit for fear and anxiety involving the extended amygdala circuit (amygdala, hippocampus, periaqueductal grey, locus coeruleus, thalamus, cingulate, and orbitofrontal areas) has lead to the hypotheses that it is this circuit which is abnormally active in panic disorder.[39]

(f) Psychological factors

Many critics disagree with the importance of biological findings in panic disorder, principally, various European workers and the cognitive theorists. They argue that panic attacks are not 'biological' and that a phobic attitude is required, and/or certain temperamental factors. Others attempt an integrated model utilizing both findings of biological differences and psychological factors of temperament and child-rearing practices.

Course and prognosis

There is limited evidence with appropriate population-based samples to clearly delineate the course of panic disorder. Available evidence suggests that panic disorder is a stable but chronic condition once criteria for the disorder are met. Most patients seeking treatment have experienced chronic, frequently chronically worsening, illness generally for 10 to 15 years prior to diagnosis.[7] However, other evidence does suggest that there is heterogeneity in terms of course.

As previously mentioned, the Klein model suggests that spontaneous panic attacks are the first manifestation of the illness, followed by anticipatory anxiety, and agoraphobia in some individuals. However, for most panic disorder patients examined closely, a panic attack is rarely the first symptom. In some studies, over 90 per cent of patients have had mild phobic or hypochondriacal, milder symptoms prior to the onset of their first panic attack.

The first panic attack is usually in a 'phobogenic' situation such as a public place, street, store, public transportation, crowd, elevator, tunnel, bridge, or open space. As mentioned, these are often preceded by stressful life events.

The earliest studies indicated low-recovery rates with chronic waxing and waning in most patients. Some individuals have outbreaks of symptomatology with less difficulty in between, but the majority of untreated individuals seem to have a more or less continuous symptom picture which ranges from mild to severe.

Recovery rates vary from 25 to 75 per cent for 1 to 2 years follow-up. Over a 5-year follow-up, only 10 to 12 per cent had fully recovered in one study and 30 per cent in another. The most common picture is about 50 per cent of patients are neither well nor very sick with mild symptoms most of the time.[40] After diagnosis and some sort of treatment, functional recovery occurs in the majority of patients.[41]

In acute pharmacological trials, 50 to 70 per cent of patients have excellent acute responses with another 20 per cent having a moderate response. With behavioural therapy, again the majority of patients recover and in some trials over 75 per cent of patients are much improved 1 to 9 years following therapy, with an average decrease in symptoms of 50 per cent.[42]

Poor responses were most consistently associated with initial high symptom severity and high agoraphobic avoidance at baseline. Poor response is also associated with low socio-economic status, less education, longer duration, limited social networks, death of a parent, divorce or unmarried status, and personality disorders.

Treatment

Introduction

Multiple effective treatments have been developed since the early 1960s and include both psychological and pharmacological treatments. Both exposure-based treatments and imipramine were shown to be effective in treating panic disorder and agoraphobia in the early 1960s.[3,4,43] Psychological-based treatments have moved increasingly towards cognitive behavioural therapy with efficacy roughly comparable to pharmacological treatments.

Imipramine and monoamine oxidase inhibitors (MAOIs) were the first medications shown to effectively treat panic disorder in the 1960s.[3,44] The high-potency benzodiazepines (alprazolam and clonazepam) were also shown to be effective in the 1980s.[7] Most now agree that the selective serotonin reuptake inhibitors (SSRIs) are the medication of first choice.[45–47]

Factors that influence the choice of initial treatment include patient preference and past history of treatment response, costs, and often availability. Medication treatment is usually easier to obtain but does involve significant costs and side effects. Although many patients prefer psychological treatments, as many as 10–30 per cent refuse treatments that involve exposure to frightening situations or resist the time and effort required. Where available, CBT can be more cost-effective than medications.

CBT has been modified in various ways to try to make it more easily available. This has included delivery in groups by computer or telephone or in shorter amounts or in a high intensity strategy with multiple hours of therapy over just a few days. All of these approaches show promise. There is clear evidence that bibliotherapy with and without phone contacts is also effective.

Eye movement desensitization and reprocessing (EMDR) was developed for treatment of post-traumatic stress disorder. The evidence available suggests that it is not effective in panic disorder.

Panic disorder can have an onset prior to adolescence, and it does occur frequently during adolescence. Although empirical data are very limited, it is generally assumed that treatments effective in adults are also effective in children. (see American Academy of Child and Adolescent Psychiatry's *Practice Parameters for the Assessment and Treatment of Children and Adolescents with Anxiety Disorders*.[48]

Medication treatments

(a) Selection of initial pharmacotherapy

Evidence indicates that medication from five classes—the SSRIs, SNRIs, benzodiazepines, tricyclics (TCAs), and monoamine oxidase inhibitors (MAOIs) are all roughly equal in their efficacy and therefore the choice of initial therapy should be made on other factors such as tolerability, cost, prior treatments, etc. (see Table 4.7.3.2). For patients with a history or concurrent depression (usually 25 per cent), the antidepressants would be preferable to the benzodiazepines. The antidepressants generally take 4 to 6 weeks to become effective, whereas the benzodiazepines begin working within the first week. There is evidence that adding benzodiazepines to the antidepressants speeds the therapeutic response.

Table 4.7.3.2 Advantages and disadvantages of various antipanic agents

	Advantages	Disadvantages
SSRIs	Well-tolerated antidepressant Safe in overdose Little weight gain Once-daily dosing	Initial activation Nausea, headache, asthenia, insomnia initially Sexual side effects
SNRIs	Very similar to SSRIs	*Hypertension*
Benzodiazepines	Rapid efficacy Reduce anticipatory anxiety Well tolerated No initial activation Safe in overdose	Sedation Some memory problems Withdrawal Abuse potential Rare sexual dysfunction
Tricyclic antidepressants	Single daily dose Less expensive Long experience Antidepressant	Initial activation Anticholinergic side effects Weight gain Orthostatic hypotension Dangerous in overdose Sexual dysfunction
MAOIs	More effective (against comorbid depression)? Antidepressant	Dietary restrictions Hypertensive crises (rare) Initial activation, insomnia Onset delayed Anticholinergic side effects Orthostatic hypotension Dangerous in overdose

A hyperstimulation reaction has been observed to all of the antidepressants and has lead to the widespread use of very low doses initially with gradual escalation. Benzodiazepines do not appear to produce the initial hyperstimulation response and are therefore preferred by many patients. However, difficulties with tapering and discontinuing benzodiazepines are the principal negative consideration for their use.

(b) Selective serotonin reuptake inhibitors (SSRIs)

In the United States, there are now six SSRIs available and three have FDA approval for panic disorder (fluoxetine, sertraline, and paroxetine—IR and CR formulations). There is no scientific or even clinical evidence to suggest significant differences in efficacy between the SSRIs in this indication. However, there are differences in side effects, principally, weight gain and discontinuation symptoms, different potential drug interactions, and availability of generic formulations.[49]

Initial jitteriness or increased anxiety are observed with the SSRIs. Therefore, treatment is often begun with the lowest doses available (see Table 4.7.3.3). Some patients do respond at lower doses, although most require higher doses (again see Table 4.7.3.3). The reason the SSRIs are generally regarded as the first choice for pharmacotherapy include their better tolerability and absence of anticholinnergic effects, compared to the TCAs. Patients often have mild difficulties with nausea, insomnia, headache. Certainly, the most problematic side effect is sexual dysfunction in both men and women, most frequently delay in orgasm. There are rare reports of extrapyramidal side effects and gastrointestinal bleeding. Discontinued too rapidly, withdrawal symptoms of headache, irritability, dizziness, can appear in the first 1 to 5 days after withdrawal, and generally clear within 1 to 2 weeks. These can generally be avoided by taper over 1 to 3 weeks.

(c) SNRIs

The SNRI venlafaxine has recently been demonstrated in large multicenter trials to be effective in the range of 75 to 225 mg/day. The side effect profile was similar in severity and symptomatology to the SSRIs, although a small number may develop sustained hypertension. There is some evidence that venlafaxine can result in higher rates of death from overdose than the SSRIs.

(d) Tricyclics (TCAs)

The first trial demonstrating that imipramine was significantly better than placebo was published in 1964[3] and was followed by multiple controlled trials demonstrating its effectiveness. There is also a significant number demonstrating effectiveness of clomipramine and some even demonstrating greater efficacy than imipramine.[50,51] Some data are supportive of desipramine and nortriptyline. Consistent problems with poorer tolerability compared to the SSRIs has lead to the TCAs being used only infrequently.

(e) Benzodiazepines (BZs)

The most widely studied and utilized benzodiazepine has been alprazolam. The largest trial was the Cross National Collaborative

Table 4.7.3.3 Medication doses in treatment of panic disorder

	Starting dose mg/day	Therapeutic range mg/day
SSRIs		
Paroxetine	10–12.5 CR	10–40*
Fluoxetine	2.5–10	10–20
Sertraline	25	50–200**
Fluovoxamine	50	100–300
Citalopram	10	20–30***
Escitalopram	5	5–10
SNRIs		
Venlafaxine	37.5	75–225
TCAs		
Imipramine	10	50–200
Clomipramine	25	25–150
BZs	TID or QID	Acute total daily dose
Alprazolam	0.25–0.5	2–10†
Clonazepam	0.25–0.5	1–4
Lorazepam	0.5	1–7
Diazepam	5	5–40
MAOIs	BID	
Phenelzine	15	15–45 (or 90)
Tranylcypromine	10	10–40 (or 70)

*40 mg demonstrated as target dose in RCT.

**All doses equivalent in one trial 50, 100, 200.

***In one trial, 20–30 was more effective than 40–60 mg/day.

†Mean dose 5.4 mg/day in largest trial.

Panic study involving more than 1000 patients and 11 trials, the majority of which were double-blind.[7] Alprazolam was effective against all of the symptoms of panic disorder, was comparable to imipramine and better tolerated. A sustained release form of alprazolam which can be taken once-daily is currently available. This formulation's long half-life appears to have solved the inter-dose rebound symptoms and 'clock watching' which was problematic with the immediate release form.

Clonazepam has also been demonstrated to be effective and is FDA approved for panic disorder in the United States.[52] Diazepam and lorazepam have also been shown to be clinically effective. Although generally well tolerated, the benzodiazepines principal side effects include sedation, occasional ataxia, slurred speech, and small increases in memory complaints. The largest concern and controversy has been with the possibility of dependency and the possibility of recreational abuse. However, the *American Psychiatric Association Task Force on Benzodiazepine Dependence, Toxicity and Abuse*, based on multiple large trials stated 'there are no data to suggest that long-term therapeutic use of benzodiazepines by patients commonly leads to dose escalation or to recreational use'.[53] Doses in long-term treatment are either similar to short-term or lower. Discontinuation symptomatology is perhaps the largest problem with the benzodiazepines. Panic patients have more difficulty discontinuing benzodiazepines than patients with generalized anxiety disorder. Symptoms are often seen during taper and are greatest during the last part of taper and the first week after taper. It is inconclusive whether these symptoms represent withdrawal, rebound or relapse, or the combination. However, it is clear that abrupt discontinuation results in greater symptomatology than a gradual taper. Most experts suggest a taper over several months (2–4 months).[54] Gradual taper and personality issues (more symptoms with a higher anxiety sensitivity and avoidance) are more critical than half-life of the medication.

Daily doses of alprazolam have varied from 2 to 10 mg. Greater efficacy is generally seen with higher doses, and the largest trial averaged 5.4 mg/day[7] (see Table 4.7.3.3).

Doses of clonazepam are often 50 per cent or less than doses with alprazolam.[52] Some clinicians prefer clonazepam over alprazolam because its longer half-life allows it to be used less frequently each day.[52] However the recent availability of an extended release once daily alprazolam (alprazolam ER) has perhaps reversed that issue. Although the BZs are generally safe in overdose, there is recent evidence that alprazolam is associated with more morbidity than the other BZs. There are studies demonstrating that lorazepam is effective averaging 7 mg/day and 5 to 40 mg/day is effective in diazepam trials (Table 4.7.3.3). The most important aspects of benzodiazepine treatments are the rapid response and increased tolerability and perhaps the greater reduction in everyday anticipatory anxiety (Table 4.7.3.2). Certainly, the largest drawbacks are in general the lack of efficacy against depression and difficulties with the tapering and discontinuation of treatment.

(f) Monoamine oxidase inhibitors (MAOIs)

Many experienced clinicians feel the MAO inhibitors may be the most effective medication class in treatment of panic disorder. However, there is only one scientifically rigorous trial supporting its use.[44] Further, concerns about its safety and the requirement of a tyramine diet limit their use. Modern studies aimed at developing reversible inhibitors of the MAOI enzyme that could eliminate the dietary restrictions have unfortunately been disappointing. The MAOI-B inhibitor selegiline has recently been made available in the United States but there are no known studies of its use in panic disorder.

Patients must also avoid sympathomimetic agents frequently found in decongestants, the antibiotic linezolid, meperidine, fentanyl, and tramadol, serotonergic agents like fenfluramine and the migraine triptan medications. Other significant side effects include weight gain, sexual dysfunctions, postual hypotension, anticholinergic side effects, and sleepiness.

(g) Other antidepressants

There are two uncontrolled trials of buproprion and buproprion SR in panic disorder, one positive and one negative. Mirtazepine has limited evidence supporting its use, as does inositol. Reboxetine has positive and negative evidence, as does with buspirone.

(h) Other agents

Although there is no evidence that conventional antipsychotic medications are effective, there is growing clinical use of the atypical antipsychotics, particularly in treatment resistant patients. There are single trials suggesting the anticonvulsants valproate and levetiracetam may be effective.

(i) Antihypertensives

Although widely used by non-psychiatric physicians in panic patients, the evidence suggests that propranolol, although it does reduce heart rate, is ineffective in treating panic disorder.

(j) Second-line medication treatments

There is limited scientific evidence to guide the clinician in the choice of the second course of treatment if the first is ineffective. Clinically, if the patient has experienced some benefit from the first medication, most clinicians would either add a benzodiazepine or CBT (which has been shown to work in SSRI failures). If the first treatment is completely ineffective, certainly switching treatment to a different SSRI or a different class of medications would be reasonable. If the second-line treatment is also ineffective, there is preliminary evidence that use of the atypical antipsychotics, olanzapine, or resperidone might be appropriate in severe non-responsive patients, and the other atypicals are also likely to work.

(k) Length of treatment

If patients do respond to an antidepressant, continuing treatment for 6 months or longer generally results in continued improvement and a decreased risk for relapse and recurrence, especially if symptoms remit.[55] Response is generally retained as long as medications are maintained. Most studies, clinical experience, and consensus opinion suggests continuation of effective medications for 12 to 18 months or longer.

(l) Discontinuing treatment after an effective response

Although antidepressants can be tapered over 10 days, clinical evidence recommends a much slower taper involving weeks to months. With benzodiazepines, it appears critical to taper even more slowly. Discontinuation symptoms are common but are significantly minimized if taper is accomplished over 2 to 4 months.[56] CBT has been demonstrated to be helpful in decreasing discontinuation difficulties when focused on the sensations, bodily symptoms, and catastrophic misinterpretations that are often seen.

Psychological treatments

(a) Psychodynamic psychotherapy

Although psychodynamic psychotherapy remains a popular treatment for all psychiatric disorders including panic disorder, there has been very little research demonstrating its efficacy in panic disorder. There is one large case-report study of patients with panic disorder reporting that most patients did respond well. One trial compared clomipramine with clomipramine plus 15 weekly sessions of brief dynamic psychotherapy. Patients in both groups responded well with 75 per cent of the patients in the clomipramine group being panic free and all of the patients in the combination group.

In an extension of this work, an emotion-focused treatment for panic disorder has been developed which explores typical fears of being abandoned or trapped as stimuli for panic attacks. This often involves a 12-session acute treatment with six sessions of monthly maintenance in which patients are encouraged to identify, reflect upon, and attempt to change problematic feelings and their responses. A specific form of dynamic psychotherapy 'panic-focused psychodynamic psychotherapy' which is generally applied in 3-month increments, was recently shown to be effective in a RCT.[57]†

Behavioural treatments

(a) Exposure treatments

Behavioural treatments which utilize *in vivo* exposure to phobic situations have been the mainstay of the behavioural treatments of panic disorder. They are based on the theory and evidence that patients who enter a feared situation experience habituation of their anxiety whether they are exposed slowly or suddenly and extensively (flooding). The critical nature of exposure for improvement was made clear in one study in which patients being treated with imipramine received no improvement from imipramine when given anti-exposure instruments.[58] However, they significantly improved if it were simply suggested that they re-expose themselves to their previously phobic situations when ready. Exposure treatment is consistently associated with long-lasting continuation of acute improvements even without formal follow-up treatment.[59]

Studies have compared use of exposure therapy to cognitive behavioural therapy, and have found them to be essentially equal in efficacy.[60] At this point, exposure has become an integral part of most CBT protocols.

(b) Cognitive behavioural therapy (see Chapter 6.3.2.1)

Cognitive behavioural therapy of panic disorder evolved from early work of Aaron Beck, but has been applied to panic disorder primarily by Barlow and colleagues working in the United States[42,61] and by Clark in the United Kingdom (see Chapter 6.3.2.1).

CBT for panic disorder usually begins with education (e.g. symptoms are part of body's fear response and aren't dangerous) about panic and the cognitive model of panic attacks, use of diaries for self-monitoring of symptoms, cognitive restructuring, habituation to fearful cues including internal cues (e.g. dizziness, tachycardia, etc.) and external situations (e.g. public places, elevators, etc.), anxiety management techniques (e.g. diaphragmatic breathing), and education to prevent relapse. Although evidence suggests that breathing retraining is not an essential component of effective CBT, it is widely utilized as an anxiety management technique.

The most informative trial to date was a multicentre (N 312) 11-week acute trial with a 6-month follow-up for responders and a 6-month follow-up after discontinuation of treatment.[62] This trial compared imipramine to cognitive behavioural therapy, their combination, and placebo. Improvement in all the active treatment cells was approximately equal at the end of the acute trial and significantly greater than placebo on most measures. This was also true at the 6-month follow-up. Interestingly, responders to imipramine had a more robust response than responders to cognitive behavioural therapy alone. The combination of imipramine plus cognitive behavioural therapy was significantly better than CBT or imipramine treatment alone at the 6-month point where all treatments were still present. However, at follow-up 6 months after treatment, none of the treatment cells were statistically different from placebo. There are many well-controlled trials demonstrating that CBT delivered in various forms and formats is effective. Its effects are robust, and at least comparable to medications.[62–64] Benefits are generally long-lasting, with or without booster sessions in follow-ups to 5 years.[63,65] CBT delivered in a group format is also effective.

Continuation/maintenance treatments

There are now a series of studies utilizing antidepressants or benzodiazepines as maintenance treatment for 6 and 12 months for panic disorder and agoraphobia. In most trials, treatment gains from acute treatment are almost always maintained, and generally are extended while the medication is continued. In a 12-month trial comparing clomipramine and placebo, the clomipramine group continued to improve and tolerated the medication well. Placebo patients who were switched to active medication matched the good responses of the clomipramine group.

In a 6-month continuation study of alprazolam patients following an 8-week initial trial, the group maintained their efficacy with a dose at the end of the 8-week trial of 5.1 ± 2.3 mg/day. This decreased to 4.7 ± 2.1 mg/day at week 32 and subsequent follow-up 1 to 2 years later found most patients' doses had drifted down to 1 to 2 mg/day.

In the large follow-up to the Phase II Cross-National Panic Trial, there was a 32-week double-blind comparison of alprazolam, imipramine, and placebo in 181 patients. Again, efficacy was maintained with both medications with no escalation of dose. Patients on both active treatments generally extended their improvement, although the placebo patients tended to lose some efficacy and certainly had a higher drop-out rate. A long-term extension continued paroxetine, clomipramine, and placebo in 176 patients following an acute trial. During the 1-year extension, both the paroxetine and clomipramine patients continued to improve and again, placebo patients tended to lose some of their initial response.

As evidence has accumulated of the high relapse rate with discontinuation of effective medication treatments for panic disorder, longer-term treatment, generally 6 to 18 months, has become routine. Although not well documented, perhaps one of the more important issues is that it appears that patients not only continue to improve for the first 6 months but that improvements continue to be extended the longer patients are on treatment, perhaps even throughout the first 2 years of treatment.

Prevention of recurrence

Available evidence remains inconclusive about the percentage of patients who will relapse if effective pharmacotherapy of panic disorder is discontinued. Early estimates suggested that most patients relapse. It remained the prevailing opinion that 35 to 85 per cent of patients relapse after antidepressants or benzodiazepines were discontinued. However, one trial reported almost no relapse after discontinuation of clomipramine patients, perhaps because they used a gradual taper.

The early trial by Zitrin *et al*. reported only 26 per cent relapse.[65] In a modern trial comparing imipramine and cognitive behavioural therapy, the imipramine relapse rate was 40 per cent. One of the most carefully performed relapse prevention trials followed a fixed-dose study of paroxetine. After acute treatment, patients were re-randomized in double-blind fashion to receive either paroxetine at their prior dose, or to placebo for an additional 3 months of treatment. Interestingly, only 30 per cent of the patients randomized to placebo relapsed, compared with 5 per cent relapse if paroxetine was continued.[47] The relapse rate after medication discontinuation in a recent trial was only 14 per cent. Although these studies certainly need replication, it suggests that the relapse rate may be lower than previously estimated if patients are slowly tapered and carefully followed.

There is one small study that suggests that the relapse rate is lower if treatment is longer. Mavissikalian followed a small group of patients who responded to imipramine, discontinuing some after 6 months of treatment and the other group after 18 months of treatment.[66] In these patients, there was an 80 per cent relapse rate in the 6-month treatment group, but only 20 per cent in the 18-month treatment group. This suggestive finding is consistent with clinical experience but certainly needs replication; however, it is certainly supportive of the general recommendation of continued treatment for 12 to 18 months if effective.

There is certainly a strong suggestion that rapid taper of benzodiazepines produces significant withdrawal symptoms which probably stimulates relapse.

Management

The suggestions for management in this section are based on evidence of the empirically based treatments in panic disorder and agoraphobia, much of which has been reviewed above. However, as in treatment of all patients, there are suggestions that also involve the 'art' of treating these patients which have evolved, but have not been empirically studied or confirmed.

Management of the uncomplicated patient

As reviewed above, it appears clear that the average patient with panic disorder can be treated with a variety of medications or exposure-based and/or cognitive behavioural treatments designed for panic disorder with approximately equal efficacy. There are some patients who have strong feelings or prejudices for or against both medication and cognitive behavioural treatments. Given that situation, as well as the lack of any clear reason to choose one treatment over the other, ethical practice would dictate offering patients a choice of treatment. There is also some evidence that patients will respond better to the treatment they 'believe in'. Unfortunately, the types of treatments are not equally available in all settings or all countries. Psychiatrists tend to use medication treatment with education, exposure, and cognitive based work of a less systematic nature than psychologists and other non-physician caregivers. Although many behaviourally and cognitively oriented psychologists do affiliate with psychiatrists and other physicians to provide medications, for many this ease of combination treatment is not readily available.

As mentioned, most psychiatrists utilize one of the medications mentioned above, supplemented by clear educational efforts with the patient and pertinent family members. This generally includes use of some written material which the patient and spouse read and discuss with the psychiatrist. (see Appendix for suggestions) Education is also a critical part of the initial treatment of patients in exposure-based treatments and cognitive behavioural therapy. These educational efforts are almost always very helpful and in the more mildly symptomatic patients may suffice. Certainly a central issue to increase the therapeutic alliance is to make clear that the therapist does understand that panic attacks involve marked 'real' physical symptoms which are extremely frightening, even though they are not dangerous and are short-lived generally.

For the patient who will be prescribed medication, it is most reasonable to offer a discussion of which medications might be appropriate, and the pros and cons of each. As outlined in Table 4.7.3.2, each medication is different, and depending upon the individual patient's needs and previous experience, any of the classes of medications might be appropriate. As mentioned, current opinion would suggest that the medication of first choice, would probably be an SSRI. In a recent meta-analysis of all the effective medications utilized in the treatment of panic disorder, the SSRIs were shown to be more effective than the other classes.[67] Coupled with their greater tolerability, lack of weight gain, and safety in overdose, they would appear to be the logical first choice.

For clinical and other practical issues, all patients should be told initially that whatever medication is the initial treatment, there are multiple other effective medications. It is important to emphasize that there is little way of knowing which specific medication is most appropriate for which patient and that the initial choice may not be effective, but subsequent choices are likely to be effective.

There are few data to direct the choice of the medications beyond those favouring the SSRIs mentioned above (see Table 4.7.3.2). There is only one trial documenting a difference in patient type leading to a choice of medication. In the large cross-national comparison of imipramine, alprazolam, and placebo, patients with predominantly respiratory symptoms responded better to imipramine.[68] Similarly, patients with a predominantly cardiovascular symptom picture responded better to alprazolam than imipramine. Otherwise, there are no data suggesting a particular medication for a specific patient beyond the various advantages and disadvantages listed in Table 4.7.3.2.

Once medication is chosen, it is prudent to begin at the lowest dose possible (see Table 4.7.3.3). This beginning low dose also extends to the benzodiazepines but is less critical since they are not associated with an initial hyperstimulation reaction. This is one of the reasons why benzodiazepines are easier to utilize and usually more popular with patients who somehow realize that from previous experience or feedback from other patients that they are not associated with an initial worsening of symptoms and are better tolerated overall. At a practical level, management of the worries about the initial hyperstimulation reaction is one of the most important issues in the psychopharmacological management of

panic disorder patients. If handled incorrectly, this issue can lead to a drop-out rate that reaches 25 to 50 per cent. With proper reassurance and close follow-up of patients, this drop-out rate can be reduced to almost zero. Patients need to be told that hyperstimulation can occur in one-third of patients but is transient and not dangerous. Because of the inherent anxiety and even phobia about taking medications, this reassurance is not usually sufficient and patients need to be invited to contact the treating physician with any anxiety or questions they might have about taking medications coupled with a quick response to their concerns.

After initial tolerance of medication is established, the dose can be raised over several weeks to a target level. Obviously, if a patient does not show a response at lower doses, the medication should be raised to maximum doses (Table 4.7.3.3).

It is important for the patient and physician alike to keep in mind that effectiveness of medications often requires a significant amount of time. The antidepressants as a class routinely take 2 to 6 weeks and with certain medications and patients as much as 6 to 12 weeks before significant effectiveness is established. This is less an issue with benzodiazepines, where initial effectiveness is generally seen in the first week or two, but there too the appropriate doses must be obtained, which often takes several weeks. Higher doses of all medicines are needed to reduce agoraphobic avoidance.

As mentioned, if agoraphobic patients do not gradually re-expose themselves to situations they fear, their avoidance fears will not be decreased. This exposure to their actual phobic situations can be accomplished in many ways. Some patients are capable of gradually re-exposing themselves after they understand the principles of exposure and the need to remain in the situation until their fears diminish. They may need help in establishing a hierarchy of their fears, although it is not actually necessary that they in fact do work-up the hierarchy in a gradually increasing fashion from 'least feared' to 'most feared'. However, it is often easier for most patients to conceptualize and accomplish it in this fashion.

Many therapists develop a hierarchical list and then monitor the patient's progress on a regular basis. Use of a standard scale which monitors the various symptom domains can be very helpful, such as *The Panic Disorder Severity scale* (PDSS).[69] There is evidence that the exposure must be regular, and extensive, often on a daily basis. Also, encouragement from the therapist and partner have both been shown to be important. Some patients can re-enter their phobic situations better if supported by their partner, or other phobics from a support group. If they are particularly afraid, an *in vivo* therapist (often recovered phobics) can be very helpful. The critical issues appear to be approaching their fears in a consistent and systematic basis in the real phobic situations accompanied by encouragement and support.

In a similar fashion, many psychiatrists combine principles of cognitive behavioural therapy without embarking on a formal cognitive behavioural therapy programme. Certainly, this should always involve education about the illness and its treatments. Other elements of identification and challenge of catastrophic thinking are widely applied by psychiatrists, but in a less systematic fashion than in formal cognitive behavioural therapy protocols.

It is important that from the beginning most patients be told that if medication treatment is effective, the expectation is to continue the medication for 12 to 18 months. An important issue to negotiate with the patient is how, when, and if effective pharmacotherapy should be tapered and discontinued.[70] Most evidence and experience suggests that patients be continued long enough to receive maximum benefit from medication treatment. In that context, patients should have experienced symptomatic and functional recovery to a maximum extent possible before discontinuation is considered. Patients should have regained a sense of confidence and control of their symptoms and lives. This might be conceptualized as a 'period of normal living' after attainment of symptomatic control before consideration is given to discontinuing an effective treatment. Relapse rates after such a remission are lower. Because the principal danger of discontinuation is relapse, the time should be carefully chosen. This should be a time when potential disruption from discontinuation symptoms and/or relapse would be least problematic.

There are strong suggestions in the literature, some of which is reviewed above, that all medications, and certainly the benzodiazepines, should be tapered very slowly, probably over 2 to 6 months, if possible.[54] This is both to minimize withdrawal symptoms which are especially frightening to panic patients and to observe for relapse symptoms as medications are slowly tapered.

The strongest reasons for discontinuing effective pharmacotherapy are the problematic side effects and expense.[70] Because this is a syndrome frequently seen in young women, the most important reason may be the wish to conceive a child or the onset of pregnancy. Certainly, routine practice is to try to taper and discontinue all medications before or during pregnancy, but sometimes this is not possible. There are now a series of women who have delivered normal children after having tried unsuccessfully to discontinue medication during pregnancy.

Many patients want to manage their own symptoms without the use of medications, and this is also a reasonable reason to taper and discontinue medications, if strongly felt by the patient.[70] The therapist should explore, however, unreasonable prejudices against the use of medications stimulated by reading, television shows, relatives, or even well-meaning physicians. Most in the field now believe that panic disorder and agoraphobia are conditions similar to hypertension and diabetes in the sense that most patients do not like the thought that they are ill and resist compliance with medication treatment. However, treatment is beneficial and not harmful, and patients often need encouragement and education in order to agree to a programme where they continue medications rather than press to discontinue them.

If medications are discontinued, the patient should be followed closely, at least by telephone, for difficulties that could include withdrawal symptoms, especially with benzodiazepines, or incipient relapse. If relapse does occur, evidence suggests that patients will respond to reinstitution of the same medication treatment regimen. If symptoms and/or functional disability associated with relapse are problematic, patients should be offered retreatment with the same medication or offered other effective non-medication treatments.

Psychological treatments

Management of patients with predominantly psychological treatments also begins with the use of educational materials, as is frequently the case with medication treatment of panic disorder. Almost all psychological treatments involve some sort of exposure-based treatment. In some, this is the predominant modality with considerable variation on how exposure to feared situations is

accomplished. Although some initial exposure therapy in particularly frightened patients may be accomplished in imagination prior to *in vivo* exposure, most exposure treatments are usually attempted *in vivo* from the outset. Most treatments have been shown to be effective if they involve *in vivo* exposure. Gradual exposure is the norm, although some programmes use very rapid exposures often called 'flooding', which involves exposure to multiple phobic situations rapidly over several days. Most programmes involve therapist-assisted exposure, sometimes utilizing professional *in vivo* therapists or volunteers who accompany phobics into their feared situations. Partners of patients are often enlisted as assistants, and there is some evidence that this more effective than non-partner exposure aides.

Most exposure-based treatments involve frequent, often daily practices involving several hours. Often the critical issue is adequate support of the patient to accomplish this much exposure 'homework', as well as encouragement and praise.

Almost anything that can help the patient accomplish the actual exposure appears to be useful and helpful. For instance, manuals and computer programs, as well as telephone-based supervision and encouragement have been shown to be effective. Although many therapists utilize relaxation techniques, applied relaxation has been the most widely utilized and effective. Many therapists also employ breathing retraining, encouraging people who hyperventilate to slow their breathing by utilizing their diaphragm. Although both have been shown to be effective and are widely utilized, other studies suggest they are not essential components of treatment, and their use does vary. Although use of benzodiazepines do decrease patient anxiety about exposure, evidence suggests that the benzodiazepines interfere with the cognitive benefits and habituation effects of exposure.

Cognitive restructuring involves the patient and therapist identifying the so-called 'automatic thoughts' they have with and after each panic attack. These are the misinterpretations that patients make about what these symptoms mean. For instance, the patient and therapist together identify that these symptoms often trigger thoughts that they are very ill, having a heart attack, or perhaps even dying. Over several sessions, these are identified and it is made clear to the patient the power these thoughts have to frighten them. At that point a number of strategies can be tried to try to correct these cognitions. Some involve attempts to compute the actual probability of the catastrophic consequences that patients fear. Others involve 'decatastrophizing' in which the ultimate consequences that patients fear are exposed, and generally can then be disavowed by the patient as extremely unrealistic. Patients can be taught to correct these thoughts or substitute more positive self-statements in their place. These skills are then worked on as 'homework', including actual exposure in which negative thoughts are identified and challenged *in vivo*.

The other usual aspect of cognitive behavioural therapy for panic disorder is interoceptive exposure, in which physical symptoms that frighten patients are identified and then they are taught to habituate to those symptoms and challenge the negative cognitions that arise with them. This can easily be accomplished in an office setting. For instance, if the patients are afraid of dizzy feelings, they can be spun in a chair. If they have fears of fast heartbeat, they can run up the stairs and challenge the negative cognitions that arise.

Both exposure-based treatments and cognitive behavioural therapy often are delivered in an 8- to 16-week treatment format, with varying frequencies of follow-up and attempts are underway to shorten these treatments. There is evidence that patients failing to respond to CBT often respond to subsequent medication treatment.

Treatment of comorbid patients

Treatment of the panic disorder patient comorbid with substance abuse is probably the most difficult challenge. In general, the substance abuse problem tends to be predominant even if it were temporally secondary to panic disorder symptoms. Therefore, treatment of substance abuse generally has to be initiated and completed first, although as soon as possible treatment of panic disorder needs be initiated.

Treatment of comorbid social phobia, obsessive–compulsive disorder, or GAD has recently been made somewhat simpler with the demonstration that the SSRIs are effective in these other conditions as well. Although not yet empirically demonstrated, it is reasonable to expect that an SSRI would effectively treat the panic disorder as well as the other comorbid anxiety disorders. This is an important area for future research. Behavioural treatments specific to obsessive–compulsive disorder and to social phobia may well be needed in addition to medication treatment.

The most common comorbidity is with depression and again one of the advantages of antidepressants is probable dual treatment of panic disorder and depression.[71] Perhaps the most important management issue is recognition that comorbid depression carries with it a marked increase in suicide risk. As mentioned, there is some evidence that depressed panic disorder patients respond better to MAOIs.

Resistance to treatment

There are relatively few systematic data about treatment options for patients resistant to initial medication or to exposure- or cognitive behavioural-based treatments. Generally, however, most patients can be tried on another medication, often with success. If they are on medicine and have not tried exposure or cognitive behavioural therapy that should definitely be added. The converse is also true. Non-response can often be traced to inadequate doses or blood levels of the medication or an inadequate length of trial. Comorbid psychiatric and particularly comorbid medical conditions need to be ruled out. Apparent resistance is often related to concomitant personality disorders or failure of agoraphobics to actually attempt exposure treatments. True resistance to one medication is sometimes overcome by a switch to another medication or to two medications at a time. If the combination of an antidepressant and benzodiazepine has not been tried, that is often the first attempted combination. In highly resistant patients, sometimes a combination of tricyclic and SSRI antidepressants or an atypical antipsychotic can be utilized.

Pregnancy

Panic disorder occurs disproportionately in young women making the issue of pregnancy a critical one. The course of panic disorder through pregnancy is highly variable. Use of SSRIs may be associated with low birth weight and a higher rate of spontaneous abortions, cardiac abnormalities, pulmonary hypertension, and withdrawal symptoms in the newborn if used late in pregnancy.

In general however, increases in congenital abnormalities have not been observed with the SSRIs. Whether benzodiazepines are associated with major malformations like cleft palate is unclear. Use of benzodiazepines near delivery is associated with sedation in the newborn. Also, both antidepressants and benzodiazepines are secreted in breast milk. For all these concerns, CBT is strongly recommended for pregnant women and should be considered in women planning pregnancy.

Ethical issues

The principal ethical issues concern the availability of treatment. Because the two types of effective treatments (medications and exposure or cognitive behavioural therapy) are not widely or equally distributed in all practices or locations, sometimes caregivers face a difficult ethical choice of having only one type of treatment available. In these instances, patients should be informed of the limitations and participate in the choices made.

Most of the treatment experience and certainly the empirical evidence has been in Caucasian patients. We do know that symptoms are different across ethnic groups and that response is often less positive in non-Caucasian groups. Treatments need to be tested and developed for all ethnic and national groups as part of the ethical development of the field.

Possibilities for prevention

The best evidence now suggests that panic disorder is often preceded by an anxiety pattern in childhood. In Kagan's model of behaviourally inhibited children, there is certainly a tendency for the pattern to persist throughout life, but some children appear to lose this trait during their development. This may well be related to parental child-rearing practices in which children are encouraged to face issues they fear rather than be withdrawn and fearful. Research is needed to explore whether different parental rearing practices or educational efforts or early treatment with these children can reduce later development of anxiety disorders. If so, these efforts at the public health and school level need to developed.

The other major preventable aetiological consideration for panic disorder and agoraphobia has been the evidence of negative traumatic events occurring in the childhood of adults with panic disorder. Preventative efforts need to be aimed at these issues through public education and education of caregivers of children. Also, one of the intervention goals in helping a traumatized child should be to prevent future development of anxiety disorders and other problems.

Further information

Barlow, D.H. and Craske, M.G. (2000). *Mastery of your anxiety and panic (MAP-3): client workbook for anxiety and panic* (3rd edn). Oxford University Press, New York.

Barlow, D.H. and Crawke, M.G. (2000). *Mastery of your anxiety and panic (MAP-3): client workbook for agoraphobia* (3rd edn). Oxford University Press, New York.

Pollard, C.A. and Zuercher-White, E. (2003). *The agoraphobia workbook: a comprehensive program to end your fear of symptom attacks.* New Harbinger, Oakland, CA.

Wilson, R.R. (2003). *Facing panic: self help for people with panic attacks*. Anxiety Disorders Association of America, Silver Spring, MD.

Zuercher-White, E. (1999). *Overcoming panic disorder and agoraphobia: client manual*. New Harbinger Publication, Oakland, CA.

Bourne, E.J. (2005). *The anxiety and phobia workbook* (4th edn). New Harbinger Publications, Oakland, CA.

Brantley, J. and Kabat-Zinn, J. (2003). *Calming your anxious mind*. New Harbinger, Oakland, CA.

Foa, E.B. and Andrews, L.W. (2006). *If your adolescent has an anxiety disorder: an essential resource for parents*. Oxford University Press, New York.

Marks, I.M. (2002). *Living with fear: understanding and coping with anxiety* (2nd edn). McGraw-Hill, New York.

Anxiety Disorders Association of America, 3730 Georgia Ave., Suite 600, Silver Spring, MD 20910, www.adaa.org

References

1. Hippocrates (1870). *On epidemics V*. Section 82 (trans. S. Farrar). Cadel, London.
2. Benedikt, M. (1870). Uber Platzschwindel. *Allgemeine Wiener Medizinische Zeitung*, **15**, 488.
3. Klein, D.F. (1964). Delineation of two drug responsive anxiety syndromes. *Psychopharmacology*, **5**, 397–408.
4. Marks, I.M. (1969). *Fears and phobias*. Heinemann, London.
5. Klerman, G.L., Weissman, M., Ovellete, R., *et al.* (1991). Panic attacks in the community: social morbidity and health care utilization. *The Journal of the American Medical Association*, **265**, 742–6.
6. Noyes, R., Jr. and Barloon, T.J. (1997). Charles Darwin and panic disorder. *The Journal of the American Medical Association*, **277**, 138–41.
7. Ballenger, J.C., Burrows, G.D., Dupont, R.L., *et al.* (1988). Aprazolam in panic disorder and agoraphobia: results from a multicenter trial. I. Efficacy in short-term treatment. *Archives of General Psychiatry*, **455**, 413–22.
8. Burns, L.E. and Thorpe, G.L. (1977). The epidemiology of fears and phobias with particular reference to the national survey of agoraphobics. *The Journal of International Medical Research*, **5**, 1–7.
9. Goldstein, A.J. and Chambless, D.L. (1978). A reanalysis of agoraphobia. *Behavior Therapy*, **9**, 47–59.
10. Beitman, B.D., Kushner, M.G., Lamerti, J.W., *et al.* (1990). Panic disorder without fear in patients with angiographically normal coronary arteries. *The Journal of Nervous and Mental Disease*, **178**, 307–12.
11. Westphal, C. (1872). Agoraphobie, eine neuropahtische Erscheinung. *Archiv für Psychiatrie und Nervenkrankheiten*, **3**, 138–61.
12. Feighner, J.P., Robins, E., Guze, S.B., *et al.* (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, **38**, 57–63.
13. Klein, D.F. (1964). Delineation of two drug responsive anxiety syndromes. *Psychopharmacology*, **5**, 397–408.
14. Ballenger, J.C. and Fyer, A.J. (1993). Examining criteria for panic disorder. *Hospital & Community Psychiatry*, **44**, 226–8.
15. Weissman, M.M. (1990). Epidemiology of panic disorder and agoraphobia. In *Frontiers of clinical neuroscience. Clinical aspects of panic disorder*, Vol. 9 (ed. J.C. Ballenger), pp. 57–65. Wiley-Liss, New York.
16. Weissman, M.M., Bland, R.C., Canino, G.J., *et al.* (1997). The cross-national epidemiology of panic disorder. *Archives of General Psychiatry*, **54**, 305–9.
17. Lépine, J.P., Wittchen, H.U., Essau, C.A., and participants of the WHO-ADAMHA CIDI Field Trials. (1993). Lifetime and current comorbidity of anxiety and affective disorders: results from the International WHO-ADAMHA CIDI Field Trials. *International Journal of Methods in Psychiatric Research*, **3**, 67–77.
18. Lydiard, R.B. and Brady, K. (1993). Association of anxiety and alcoholism. *The Psychiatric Quarterly*, **64**, 135–49.

19. Sartorius, N., Uestuen, B., Costa e Silva, J.A., *et al.* (1993). An international study of psychological problems in primary care: preliminary report from the World Health Organization collaborative project on psychological problems in general health care. *Archives of General Psychiatry*, **50**, 819–24.
20. Kroenke, K. and Mangelsdorff, A.D. (1989). Common symptoms in ambulatory care: incidence, evaluation, therapy and outcome. *The American Journal of Medicine*, **86**, 262–6.
21. Simon, G.E. and Von Korff, M. (1991). Somatization and psychiatric disorder in the Epidemiologic Catchment Area study. *The American Journal of Psychiatry*, **148**, 1494–500.
22. Katon, W. (1984). Panic disorder and somatization. Review of 55 cases. *The American Journal of Medicine*, **77**, 101–8.
23. Ballenger, J.C. (1998). Panic disorder in primary care and general medicine. In *Panic disorder and its treatment* (eds. J. Rosenbaum and M. Pollack), pp. 1–36. Dekker, New York.
24. Goodwin, R.D., Lieb, R., Hoefler, M., *et al.* (2004). Panic attack as a risk factor for severe psychopathology. *The American Journal of Psychiatry*, **161**, 2207–14.
25. Kessler, R.C., Chiu, W.T., Jim, R., *et al.* (2006). The epidemiology of panic attacks, panic disorder, and agoraphobia in the national comorbidity survery replication. *Archives of General Psychiatry*, **63**, 415–24.
26. Torgersen, S. (1983). Genetic factors in anxiety disorders. *Archives of General Psychiatry*, **40**, 1085–90.
27. Kendler, K.S., Neale, M.C., Kessler, R.C., *et al.* (1993). Panic disorder in women: a population-based twin study. *Psychological Medicine*, **40**, 397–406.
28. Skre, I., Onstad, S., Torgersen, S., *et al.* (1993). A twin study of DSM-III-R anxiety disorders. *Acta Psychiatrica Scandinavica*, **88**, 85–92.
29. Arnold, P.D., Zai, G., and Richter, M.A. (2004). Genetics of anxiety disorders. *Current Psychiatry Reports*, **6**, 243–54.
30. Roy-Byrne, P.P., Craske, M.G., and Stein, M.G. (2006). Panic disorder. *Lancet*, **368**, 1023–32.
31. Kagan, J., Reznick, J.S., Clarke, C., *et al.* (1984). Behavioral inhibition to the unfamiliar. *Child Development*, **55**, 2212–25.
32. Rosenbaum, J.F., Biederman, J., Gersten, M., *et al.* (1988). Behavioral inhibition in children of parents with panic disorder and agoraphobia: a controlled study. *Archives of General Psychiatry*, **45**, 463–70.
33. Scocco, P., Barbieri, I., and Frank, E. (2007). Interpersonal problem areas and onset of panic disorder. *Psychopathology*, **40**, 8–13.
34. Laraia, M.T., Stuart, G.W., Frye, L., *et al.* (1994). Childhood environment of women with panic disorder and agoraphobia. *Journal of Anxiety Disorders*, **8**, 1–17.
35. Saunders, B.E., Villeponteaux, L.A., Lipovsky, J.A., *et al.* (1992). Child sexual assault as a risk factor for mental disorders among women: a community survey. *Journal of Interpersonal Violence*, **7**, 189–204.
36. Johnson, P.L., Lightman, S.L., and Lowry, C.A. (2004). A functional subset of sertonergic neurons in the rat ventrolateral periaqueductal gray implicated in the inhibition of sympathoexcitation and panic. *Annals of the New York Academy of Sciences*, **1018**, 58–64.
37. Neumeister, A., Bain, E., Nugent, A.C., *et al.* (2004). Reduced serotonin type 1_A receptor binding in panic disorder. *The Journal of Neuroscience*, **24**, 589–91.
38. Maron, E., Kuikka, J.T., Shlik, J., *et al.* (2004). Reduced brain serotonin transporter binding in patients with panic disorder. *Psychiatry Research*, **132**, 173–81.
39. Bremner, J.D. (2004). Brain imaging in anxiety disorders. *Expert Review of Neurotherapeutics*, **4**, 275–84.
40. Katschnig, H., Amering, M., Stolk, J.M., *et al.* (1996). Predictors of quality of life in a long-term followup study in panic disorder patients after a clinical drug trial. *Psychopharmacology Bulletin*, **32**, 149–55.
41. Mavissakalian, M.R. and Prien, R.F. (eds.) (1996). *Long-term treatments of anxiety disorders*. American Psychiatric Press, Washington, DC.
42. Landon, T.M. and Barlow, D.H. (2004). Cognitive-behavioral treatment for panic disorder: current status. *Journal of Psychiatric Practice*, **10**, 211–26.
43. Marks, I.M. (1987). *Fears, phobias, and rituals.* Oxford University Press, New York.
44. Sheehan, D., Ballenger, J.C., and Jacobsen, G. (1980). Treatment of endogenous anxiety with phobic hysterical and hypochondriacal symptoms. *Archives of General Psychiatry*, **37**, 51–9.
45. Jobson, K.O. and Poter, W.Z. (1995). International psychopharmacology algorithim project report. *Psychopharmacology Bulletin*, **31**, 457–507.
46. Ballenger, J.C., Davidson, J.R., Lécrubier, Y., *et al.* (1998). Consensus statement on panic disorder from the International Consensus Group on depression and anxiety. *The Journal of Clinical Psychiatry*, **59**, 47–54.
47. Ballenger, J. (1999). Selective serotonin reuptake inhibitors (SSRIs) in panic disorder. In *Panic disorder: clinical diagnosis, management and mechanisms* (eds. D. Nutt, J. Ballenger, and J.P. Lépine), pp. 159–78. Dunitz, London.
48. Connolly, S.D., Bernstein, G.A., and Work Group on Quality Issues (2007). Practice parameters for the assessment and treatment of children and adolescents with anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **46**(2), 267–83.
49. Fava, M. (2006). Prospective studies of adverse events related to antidepressant discontinuation. *The Journal of Clinical Psychiatry*, **67**(Suppl. 4), 14–21.
50. Cassano, G.B., Petracci, A., Perugi, G., *et al.* (1998). Clomipramine for panic disorder. I. The first 10 weeks of a long-term comparison with imipramine. *Journal of Affective Disorders*, **14**, 123–7.
51. Modigh, L., Westberg, P., and Eriksson, E. (1992). Superiority of clomipramine over imipramine in the treatment of panic disorder: a placebo-controlled trial. *Journal of Clinical Psychopharmacology*, **51**, 53–8.
52. Herman, J.B., Rosenbaum, J.F., and Brotman, A.W. (1987). The alprazolam to clonazepam switch for the treatment of panic disorder. *Journal of Clinical Psychopharmacology*, **7**, 175–8.
53. American Psychiatric Association. (1990). Benzodiazepines dependence, toxicity, and abuse: a task force report of the American Psychiatric Association. APA, Washington, DC.
54. Dupont, R.L., Swinson, R.P., Ballenger, J.C., *et al.* (1992). Discontinuation of alprazolam after long-term treatment of panic-related disorders. *Journal of Clinical Psychopharmacology*, **12**, 352–4.
55. Lecrubier, Y. and Judge, R. (1997). Long-term evaluation of paroxetine, clomipramine and placebo in panic disorder. Collaborative Paroxetine Panic Study Investigators. *Acta Psychiatrica Scandinavica*, **95**, 153–60.
56. Ballenger, J.C., Pecknold, J., Rickels, K., *et al.* (1993). Medication discontinuation in panic disorder. *The Journal of Clinical Psychiatry*, **54**, 15–21.
57. Milrod, B., Leon, A.C., Busch, F., *et al.* (2007). A randomized controlled clinical trial of psychoanalyltic psychotherapy for panic disorder. *The American Journal of Psychiatry*, **164**, 265–72.
58. Mavissakalian, M. (1990). Sequential combination of imipramine and behavioral instructions in the treatment of panic disorder with agoraphobia. *Archives of General Psychiatry*, **46**, 127–31.
59. Fava, G.A., Rafanelli, C., Grandi, S., *et al.* (2001). Long-term outcome of panic disorder with agoraphobia treated by exposure. *Psychological Medicine*, **31**, 891–8.
60. Ost, L.G., Thulin, U., and Tamnero, J. (2004). Cognitive behavior therapy vs. exposure *in vivo* in the treatment of panic disorder with agoraphobia (corrected from agrophobia). *Behaviour Research and Therapy*, **42**, 1105–27.
61. Barlow, D.H., Craske, M.G., Cerny, J.A., *et al.* (1989). Behavioral treatment of panic disorder. *Behavior Therapy*, **20**, 261–82.

62. Barlow, D.H., Gorman, J.M., Shear, M., *et al.* (2000). Cognitive-behavioral therapy, imipramine, or their combination for panic disorder: a randomized controlled trial. *The Journal of the American Medical Association*. **283**, 2529–36.
63. Craske, M.G., Brown, T.A., and Barlow, D.H. (1991). Behavioral treatment of panic disorder: a two-year follow-up. *Behavior Therapy*, **22**, 289–304.
64. Butler, A.C., Chapman, J.E., Forman, E.M., *et al.* (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clinical Psychology Review*, **26**, 17–31.
65. Zitrin, C.M., Klein, D.F, Woerner, M.G., *et al.* (1983). Treatment of phobias. I. Comparison of imipramine hydrochloride and placebo. *Archives of General Psychiatry*, **46**, 127–31.
66. Mavissakalian, M. (1990). Sequential combination of imipramine and behavioral instructions in the treatment of panic disorder with agoraphobia. *Archives of General Psychiatry*, **46**, 127–31.
67. Boyer, W. (1995). Serotonin uptake inhibitors are superior to imipramine and alprazolam in alleviating panic attacks: a meta-analysis. *International Clinical Psychopharmacology*, **10**, 45–9.
68. Briggs, A.C., Stretch, D.D., and Brandon, S. (1993). Subtyping of panic disorder by symptom profile. *The British Journal of Psychiatry*, **163**, 201–9.
69. Shear, M.K., Brown, T.A., Barlow, D.H., *et al.* (1997). Multicenter collaborative panic disorder severity scale. *The American Journal of Psychiatry*, **154**, 1571–5.
70. Ballenger, J.C. (1992). Medication discontinuation in panic disorder. *The Journal of Clinical Psychiatry*, **53**, 26–31.
71. Ballenger, J.C. (1988). Comorbidity of panic and depression: implications for clinical management. *International Journal of Clinical Psychopharmacology*, **13**(Suppl. 4), S13–17.

4.8

Obsessive–compulsive disorder

Joseph Zohar, Leah Fostick, and Elizabeth Juven-Wetzler

Introduction

Obsessive–compulsive disorder (OCD) is a common, chronic, and disabling disorder marked by obsessions and/or compulsions that are egodystonic and cause significant distress to the patients and their families. During the last 25 years, there has been a resurgence of studies into various aspects of OCD, including epidemiological, pathophysiological, and pharmacological investigations. With the progress in finding effective treatments for OCD, different algorithms for the management of these patients have been developed. The progress in OCD includes advanced methodologies of imaging studies (both before and after treatment), along with insight into the neurological aspects of OCD and OCD-related conditions, leading to selective treatments.

Up to the early 1980s, OCD was considered a rather rare, treatment-refractory, and chronic condition of psychological origin. Dynamic psychotherapy was of little benefit and several pharmacological treatments were attempted without much success.[1] Since then, several researchers have reported that the prevalence of OCD is around 2 per cent in the general population.[2,3] In addition, numerous studies have reported on the efficacy of various serotonin reuptake inhibitors, and consequently an understanding of the biological basis of OCD has begun to unfold.

The observation that clomipramine, a tricyclic antidepressant with a serotonergic profile, is effective in treating symptoms of OCD[4,5] has increased interest in OCD in general and in the relationship between serotonin and OCD in particular. Substantial evidence currently suggests that OCD is almost unique among psychiatric disorders, as only serotonergic medications appear to be effective in this disorder.[6] For example, non-serotonergic drugs, such as desipramine, a potent antidepressant and antipanic agent, are entirely ineffective in OCD.[7–9] This specific response to serotonergic drugs has paved the way for further research on the role of serotonin in the pathogenesis of OCD in particular, and in OCD-related disorders in general.

Epidemiology

The lifetime prevalence of OCD in the general population is between 2 and 3 per cent (i.e. it is more prevalent than schizophrenia).[2,10] This rate has been confirmed across different cultures.[3] The prevalence of OCD among children and adolescents appears to be as high as among adults.[11] However, Nelson and Rice[12] and Stein *et al.*[13] have suggested that the diagnosis of OCD by the Diagnostic Interview Schedule administered by lay people leads to overdiagnosis, and so have proposed lower prevalence rates of 1 to 2 per cent.

Men and women are equally likely to be affected, although some reports have suggested a slight female predominance.[3] During adolescence, boys are more commonly affected than girls. The mean age of onset is about 20 years of age. Single people are more commonly affected, probably representing the difficulty for people with OCD to maintain a relationship.

Patients with OCD are commonly afflicted by other mental disorders; for instance, the lifetime prevalence for a major depressive episode in these patients is around 67 per cent.[3,14] Other common comorbid psychiatric diagnoses include alcohol-use disorders, social phobia, specific phobia, panic disorder, eating disorders, and post-traumatic stress disorder (PTSD).[15] The comorbidity with schizophrenia and with tic disorders raises interesting pathophysiological and therapeutic implications. The rate of tic disorders approaches 40 per cent in juvenile OCD, and there is an increase in the prevalence of Tourette's syndrome among the relatives of OCD patients.[16]

The relationship between OCD and obsessive–compulsive personality disorder (OCPD) has been a focus of debate. Although prospective research is lacking, it appears that OCPD is not a prominent risk factor for developing OCD, as the prevalence of OCPD among patients with OCD is not far from its prevalence in other psychiatric disorders.

Clinical features and diagnosis

The diagnosis of OCD according to DSM-IV criteria is based on the presence of either obsessions or compulsions, which cause marked distress, are time-consuming (more than an hour per day), or significantly interfere with the person's normal routine and social and occupational activities. It stipulates that, at some point during the course of the disorder, but not necessarily during the current episode, the person has recognized that the obsessions or compulsions are excessive or unreasonable. However, if the patient does not recognize for most of the time during the current episode that the obsessions and compulsions are excessive or unreasonable, the diagnosis is OCD with poor insight.

If another Axis I disorder is present, it is mandatory that the content of the obsessions or compulsions is not restricted to it (e.g. a preoccupation with food or weight in eating disorders, or guilt feelings in the presence of a major depressive episode). The disturbance should not be due to the direct effects of a substance (e.g. of a drug abuse or a medication) or a general medical condition.

The obsessions are recurrent, intrusive, and distressing thoughts, images, or impulses, whereas the compulsions are repetitive, seemingly purposeful, behaviours that a person feels driven to perform. Obsessions are usually unpleasant and increase a person's anxiety, whereas carrying out compulsions reduces anxiety. Resistance to carrying out a compulsion results in increased anxiety. The patient usually realizes that the obsessions are irrational and experiences both the obsession and the compulsion as egodystonic.

Patients with both obsessions and compulsions constitute at least 75 per cent of the affected patients, with most patients presenting with multiple obsessions and compulsions. The symptoms may shift, for example a patient who had washing rituals during childhood may present with checking rituals as an adult.

OCD can express itself in many different symptoms, but the classical presentations include washing, checking, aggressive, religious, or sexual obsessions, and ordering, counting, hoarding, and symmetry compulsions. Dimensional approaches have been used to analyze these characteristic subtypes, and present the different symptoms in an innovative way.[17]

The most common pattern is an obsession with dirt or germs, followed by washing or avoiding presumably contaminated objects (doorknobs, electrical switches, newspapers, people's hands, telephones). Because it is hard to avoid the feared object is (e.g. faeces, urine, dust, or germs), patients wash their hands excessively and sometimes avoid leaving home because of their fear of germs. A second common pattern is an obsession of doubt, followed by a compulsion of checking. The person checks whether the oven is turned off or the front and back doors are closed—the checking may involve many trips back home to recheck what had already been checked. In OCD, the checking, instead of resolving uncertainty, often contributes to even greater doubt, which leads to further checking. The patients exhibit obsessional self-doubt, and feel guilty for having committed some damage (for instance, a fear of hurting someone while driving, leading to driving back over the same spot again and again). Other patterns include hoarding and religious obsessions. More recently, a dimensional approach to OCD has been launched by Leckman and colleagues, stating four symptom dimension of OCD: obsessions/checking, symmetry/ordering, contamination/cleaning, and hoarding.[17,18] Another pattern of OCD involves intrusive obsessional thoughts without a compulsion. Such obsessions are usually repetitious thoughts of some sexual or aggressive act that is reprehensible to the patient. Still another pattern is the need for symmetry or precision, which leads to a compulsion of slowness. Patients can take hours to eat a meal or shave, in an attempt to do things 'just right'. Unlike other patients with OCD, these patients usually do not resist their symptoms.

The gap between the knowledge that the symptoms are irrational on one hand and the overwhelming urge to perform them on the other hand contributes to the immense suffering associated with OCD.

OCD and schizophrenia

About 25 per cent of patients with chronic schizophrenia may also present with OCD symptoms (range 5 to 45 per cent)[19]; and 15 per cent of the patients with schizophrenia may fully qualify for the diagnosis of OCD. As in OCD, the OC symptoms in these patients will not necessarily surface unless specific questions are asked. Many patients with schizophrenia can distinguish the egodystonic, obsessive–compulsive symptoms, perceived as coming from within, from the egosyntonic delusions perceived as introduced from the outside. Follow-up studies demonstrate a diagnostic stability over the years, and it seems that the presence of OCD in schizophrenia predicts a poor prognosis.[19] Several studies among patients with schizophrenia and OCD reported an improvement in OCD symptomatology after the addition of serotonin reuptake inhibitors.[19]

The poor prognosis of patients with schizophrenia and OCD, preliminary data regarding their response to the unique combination of antipsychotic and anti-obsessive medications, along with the high prevalence of this presentation has led several researchers to suggest that a 'schizo-obsessive' category may be of value.[20]

Differential diagnosis

Personal distress and functional impairment, which are required for the diagnosis, differentiate OCD from ordinary or mildly excessive worries, thoughts, and habits. The medical differential diagnosis includes tic disorders (especially Tourette's syndrome), temporal-lobe epilepsy, trauma, and postencephalitic complications.

Psychiatric diagnoses that should be ruled out include depressive disorder, schizophrenia, OCPD, PTSD, phobias, delusions, hypochondriasis, and paraphilias. OCD can usually be differentiated from schizophrenia by the absence of other schizophrenic symptoms and by the patients' insight into their disorder. Moreover, patients with OCD usually attempt to resist the obsessions. OCPD does not have the degree of functional impairment characteristic of OCD and it is egosyntonic.

Phobias are distinguished by the absence of a relationship between the obsessive thoughts and the compulsions. The fears in OCD usually involve harm to others rather than harm to oneself. In addition, in OCD, when patients are 'phobic' they are usually afraid of an unavoidable stimulus (for instance, viruses, germs, or dirt) as opposed to the classic phobic objects like tunnels, bridges, or crowds.

Major depressive disorder (MDD) can sometimes be associated with obsessive ideas, but patients with OCD usually fail to meet all the criteria of MDD. Other psychiatric diagnoses closely related to OCD are hypochondriasis, body dysmorphic disorder, and trichotillomania. As these patients have repetitive worries or behaviours, although they are focal, they are still related to the 'OCD Spectrum'.

Course and prognosis

Many patients with OCD may have an onset of symptoms after a stressful event (e.g. pregnancy, a loss, or a sexual problem). Owing to the secretive nature of the disorder, there is often a delay of 5 to 10 years before patients come to psychiatric attention. However, the delay may shorten due to increased public awareness to the disorder through articles, books, and movies. The course of OCD is

usually long, but variable; some patients experience a fluctuating course, while others experience a chronic course.[21]

About 20 to 30 per cent of the patients show a significant improvement in their symptoms, and 40 to 50 per cent a moderate improvement. The remaining 20 to 40 per cent become chronic or their symptoms worsen.

OCD patients are prone to depression and sometimes even to suicide. A poor prognosis is indicated by yielding (rather than resisting) to compulsions, a early onset, male gender, tic related forms of OCD with associated to hoarding/symmetry compulsions, the need for hospitalization, psychotic features, a coexisting major depressive disorder, delusional beliefs, the presence of overvalued ideas (i.e. some acceptance of the obsessions and compulsions), and the presence of personality disorder (especially schizotypal personality disorder).[22–24] A good prognosis is indicated by good social and occupational adjustment and less avoidance.[21] The obsessional content does not seem to be related to the prognosis, except for hoarding, which is usually considered to have a less favourable outcome.

Aetiology

Neurotransmitters

Many clinical trials of various serotonergic drugs lend support to the hypothesis that a dysregulation of serotonin is involved in the beneficial therapeutic effect in OCD. However, this does not necessarily reflect on pathogenesis. Abnormality of the serotonergic system, and particularly the hypersensitivity of postsynaptic 5-HT receptors, constitutes the leading hypothesis for the underlying pathophysiology of OCD.[7,25–38] However, a potential role for dopamine has been emerging as well.[39]

Clinical studies have assayed cerebrospinal levels of serotonin metabolites (e.g. 5-hydroxyindoleacetic acid [5-HIAA] a 5-HT metabolite that serves as an index of 5-HT turnover)[25,26] and affinities of imipramine and paroxetine[40] binding sites on platelets show that it binds to serotonin reuptake sites,[27–30] in some studies of OCD patients. A study supporting the relationship between a decreased function of the serotonergic system and a positive response to selective serotonin reuptake inhibitors (SSRIs), demonstrated normalization of the number of platelet 5-HT transporters following treatment with different SSRIs.[31] In an earlier study, patients who responded to clomipramine had higher pretreatment levels of 5-HIAA than the non-responders.[25] Moreover, the clinical improvement was positively correlated with a decrease in the concentration of 5-HIAA in cerebrospinal fluid.[25]

Another approach is to examine peripheral measures of serotonergic and noradrenergic function in patients with OCD. In one study, clinical improvement during clomipramine therapy closely correlated with pretreatment platelet serotonin concentration and monoamine oxidase activity, as well as with the decrease in both measures during clomipramine administration.[32] Moreover, only the plasma levels of clomipramine (a potent 5-HT reuptake inhibitor), but not the plasma levels of its primary metabolite, desmethyl clomipramine (which has noradrenergic properties), correlated significantly with an improvement in OCD symptoms. These findings suggest that the effects of anti-obsessive medications, clomipramine in this study, on serotonin function are pertinent to the anti-obsessional action observed.

Additional support for the importance of serotonin in the therapeutic response to serotonin reuptake inhibitors (SRIs) in OCD came from a study by Benkelfat *et al.*[38] in which the investigators administered the serotonin receptor antagonist metergoline and placebo to 10 patients with OCD in a double-blind crossover study. Patients receiving clomipramine on a long-term basis responded with greater anxiety to a 4-day administration of metergoline when compared with the placebo phase of the study.

Additional evidence for disturbances of the serotonergic system in OCD was provided by challenge studies. Challenges with L-tryptophan,[33] *m*-chlorophenylpiperazine (mCPP),[7,34] sumatriptan (a 5-HT1D agonist[6]), ipsapirone (a 5-HT1A receptor ligand[35]), and MK-212 (a 5-HT1A and 5-HT2C agonist[36]), among others, were used to evaluate whether they worsen obsessive–compulsive symptoms or whether they elicit different physiological responses (thermal or neuroendocrine) in patients with OCD compared with controls. Only two compounds (*m*-chlorophenylpiperazine and sumatriptan) have shown behavioural hypersensitivity and neuroendocrine hyposensitivity to be characteristic of serotonergic challenges in patients with OCD. These studies may have the potential to pinpoint the receptor subtype involved in OCD, raising the possibility that 5-HT_{1B} (but not 5-HT_{1A}) could be involved in OCD.[37]

Dopamine

The most compelling evidence for dopaminergic involvement in OCD comes from the abundance of OCD symptoms in basal ganglia disorders, such as Tourette's syndrome, Sydenham's chorea, and postencephalitic parkinsonism. The therapeutic benefits obtained with the coadministration of dopamine blockers and SRIs in a subset of patients with both OCD and tic disorder[41] has also suggested a role for dopamine dysfunction. A study evaluating levels of platelet sulphotransferase, an enzyme involved in the catabolism of catecholamines (providing a marker of presynaptic dopamine function), reported a decreased level of platelet [^{3}H]imipramine binding and a parallel increase in the level of sulphotransferase activity in OCD compared with controls. This provides further support for the hypothesis of reduced 5-HT activity and increased dopamine transmission in OCD.[28,39]

Immune factors

Study of autoimmune factors has been prompted by the association of OCD and the autoimmune disease of the basal ganglia, Sydenham's chorea. This complication of rheumatic fever is accompanied by obsessive–compulsive symptoms in over 70 per cent of cases[42]: 10 out of 11 children had antibodies directed against the caudate.[42] These children had a history of obsessive–compulsive symptoms, which started prior to the onset of the chorea, reached a peak in line with the motor symptoms, and declined with their resolution. This is consistent with the hypothesis of basal ganglia dysfunction in OCD.

Antibodies against two peptides of the basal ganglia have also been found.[43] A strong connection was reported between OCD/Tourette's syndrome and the B-cell antibody D8/17, which is another antibrain antibody.[44] The specificity of these antibodies to OCD, as well as the generalizability of these rare cases, is as yet unclear.

Brain imaging studies

The use of positron emission tomography has demonstrated the presence of increased activity (i.e. metabolism and blood flow) in the frontal lobes, the basal ganglia (especially the caudate nucleus), and the cingulum of patients with OCD.[45] Pharmacological and behavioural treatments reportedly reverse those abnormalities.[46] The data from functional imaging studies are consistent with the data from structural brain imaging studies. Both CT and magnetic resonance imaging studies have found decreased sizes of caudate bilaterally. Both functional and structural imaging procedures are also consistent with the observation that neurological procedures involving the cingulum are sometimes effective in the treatment of patients with OCD.

Overall, the brain imaging research suggests a role for the prefrontal cortex-basal ganglia thalamic circuitry. Dysfunction of these circuits can be explored by neuropsychological testing and recording evoked potentials. Indeed, a study of patients with OCD demonstrated that they are slower in performing tasks involving frontocortical systems, suggesting alterations at this level.[47] An evoked potential study showed enhanced processing negativity in the frontal cortex consistent with the prefrontal hyperactivity shown in brain imaging studies.[48] Moreover, the reflection of behavioural challenge on brain activity (brain responsivity) may be a potential tool for predicting a response to successful intervention with SSRI.[49]

Genetics

A significantly higher concordance rate was found for monozygotic twins than for dizygotic twins.[50] Of the first-degree relatives of patients with childhood-onset OCD, 35 per cent are also afflicted with the disorder.[51] Although this high rate is possibly related to the early-onset subtype, it nevertheless suggests a genetic component in OCD. Genetic research has yet to find abnormalities at the 5-HT transporter gene level. A study exploring the polymorphism of the promoter region of the gene for the 5-HT transporter failed to identify any differences between patients with OCD and controls.[52] However, several studies found polymorphism of $5HT_{1B\beta}$ in OCD,[53,54] hence providing further support for the $5HT_{1B}$ involvement in OCD.

Other biological data

Sleep electroencephalography and neuroendocrine studies have found abnormalities similar to those seen in depression, such as decreased rapid eye movement latency, non-suppression on the dexamethasone suppression test, and decreased growth hormone secretion with clonidine infusions.[55,56]

Behavioural factors

According to the learning theory, obsessions can be considered conditioned stimuli. When a relatively neutral stimulus is coupled with an anxiety-provoking stimulus, through conditioning, it will produce anxiety even when presented alone. In this regards, even the thought of the anxiety-provoking stimulus can cause anxiety, similarly to Pavlov's dog, which salivated even before he actually had food. Consequently, avoidant behaviour is being adopted in order to avoid the anxiety-provoking stimulus and any other stimuli, which remind it. The compulsions are learnt as a way to reduce anxiety. Once producing a relief of the anxiety, the relief serves as reinforce to the compulsion, which are then being repeated by the patient. Through the process of conditioning, reward and reinforcement, rituals, and avoidant strategies are become fixed.

Psychological factors

The dynamic aspects of OCD were first described by Sigmund Freud, who coined the term 'obsessional neurosis'. The disorder was thought to result from a regression from the Oedipal phase to the anal phase, with its characteristic ambivalence. The coexistence of hatred and love towards the same person leaves the patient paralyzed with doubt and indecision. Freud originally suggested that obsessive symptoms result from unconscious impulses of an aggressive or sexual nature. These impulses cause extreme anxiety, which is avoided by the defence mechanisms. One of the striking features of patients with OCD is the degree to which they are preoccupied with aggression or cleanliness (anal phase), either overtly in the content of their symptoms or in the underlying associations.

Freud described three major psychological defence mechanisms that are important in OCD: isolation, undoing, and reaction formation. According to the psychoanalytical formulation, OCD develops when these defences fail to contain the anxiety. Isolation is the separation of the idea and the affect that it arouses. Undoing is a secondary defence to combat the impulse and quit the anxiety that its imminent eruption into consciousness arouses. Undoing is a compulsive act, performed to prevent or undo the results that the patient irrationally anticipates from a frightening obsessional thought or impulse. Reaction formation is related to the production of character traits rather than symptom formation (characteristic of the above defences). The trait seems highly exaggerated and inappropriate (i.e. the switch of anger and hate into exaggerated love and dedication).

Summary

The efficacy of the SRIs for OCD, together with the lack of efficacy of adrenergic antidepressants, has suggested that serotonin is involved in the pathophysiology of OCD. This relationship was validated by research on serotonergic markers in OCD and by the challenge paradigm.[6] Which type of serotonergic receptor is involved in the pathogenesis and/or the mechanism of action of anti-obsessional drugs, is still unclear. However, the possible role of $5HT_{1B}$ has emerged. Further studies are crucial for elucidating the role of serotonin and other neurotransmitters (i.e. dopamine) in the pathophysiology and management of OCD.

The pharmacological treatment of OCD

Since the early 1980s, several potent SRIs have been studied extensively in OCD. Aggregate statistics for all SRIs suggest that 70 per cent of treatment-naive patients will improve at least moderately.[57]

Efficacy of serotonergic versus adrenergic antidepressants

Whilst anecdotal reports have suggested that clinical benefit can be obtained with a range of reuptake blockers, effectiveness has only been demonstrated consistently for the SRIs. Several studies have directly compared clomipramine with other antidepressants with a consistent finding: antidepressant drugs that are less potent SRIs than clomipramine are generally ineffective in OCD.[7–9,25]

In the late 1960s, clomipramine was the first reported effective medication for OCD.[4,5] Since then, numerous placebo-controlled studies have clearly shown clomipramine's effectiveness, and this has been confirmed in a United States multicentre controlled trial ($n = 520$).[58] In this study, after 10 weeks of treatment, 58 per cent of patients treated with clomipramine rated themselves much or very much improved versus 3 per cent of placebo-treated patients.

Besides the SRI clomipramine, the newer non-tricyclic SSRIs, such as fluoxetine, fluvoxamine, paroxetine, sertraline, citalopram, and escitalopram are gaining acceptance as effective alternatives for the treatment of OCD in controlled studies. Actually, they were found to be as effective as clomipramine.[59–62] Since SSRIs are less toxic in case of overdose, and as they have less cholinergic side-effects, they are considered as a first-line treatment for OCD.

(a) Onset of treatment response

It has been suggested that a relatively long period, up to 8 or even 12 weeks, is needed before one can consider a serotonin reuptake inhibitor to be ineffective. Several months' treatment is often needed to achieve a maximum response.

(b) Long-term treatment

Most patients relapse after prematurely discontinuing treatment, but, as stated above, it may take many months for a maximum response to be seen. Pato *et al.*[63] reported that 16 out of 18 patients with OCD relapsed within 7 weeks after stopping clomipramine, although some had been treated for more than a year (mean, 10.7 months). All patients regained the therapeutic effects when clomipramine was reintroduced. Leonard *et al.*[9] examined the effect of clomipramine substitution during a long-term clomipramine treatment in 26 children and adolescents with OCD (mean duration of treatment was 17 months). Half the patients were blindly assigned to 2 months of desipramine treatment, and then clomipramine was reintroduced. Almost 90 per cent relapsed during the 2 months' substitution period compared with only 18 per cent of those kept on clomipramine throughout the study. Therefore, it seems advisable that patients with OCD should be maintained on anti-obsessive medications for a long period, and certainly for more than a year before a very gradual attempt is made to discontinue the treatment.

The maintenance dose needed in OCD is also unclear. In a study that examined this issue, Mundo *et al.*[64] investigated the effect of dose reduction in patients previously treated successfully with fluoxetine. Patients were randomized to receive the same drug dosage or to receive a reduced dose. It appears that 'the dose that makes you well keeps you well'; i.e. that medium to high doses of SRIs are required.

(c) Drug dosage

Higher doses of SSRIs have been used in the treatment of OCD as compared to treatment of depression. Two fixed-dose studies using fluoxetine and one pan-European study with paroxetine have found some advantage with using higher doses, and those effects were found with citalopram and escitalopram studies.[6,37,62,65–67] A theoretical basis for this clinical finding, which is related to the 'stickiness' of the $5HT_{1B}$ receptor has been reported.[67,68]

Comparative studies of clomipramine versus SSRIs

The introduction of SSRIs has raised the question regarding the comparative efficacy of clomipramine versus that of the SSRIs. SSRIs are important alternatives to clomipramine, since their range of side-effects is different (absence of anticholinergic side-effects, sedation, safety with overdose, etc.). Although SSRIs may be associated with sexual side-effects, headaches, and appetite disturbances, these side-effects are usually less troublesome as compared to clomipramine's side-effects.

Fluoxetine was compared with clomipramine in 11 patients with OCD in a 10-week crossover study.[69] Although no significant differences were noted regarding clinical efficacy, the proportion of fluoxetine non-responders who later responded to clomipramine tended to be higher compared with the clomipramine non-responders who were switched to fluoxetine. However, patients reported significantly fewer side-effects while on fluoxetine. Freeman *et al.*[70] compared the efficacy of fluvoxamine and clomipramine in a multicentre randomized double-blind parallel-group comparison in 66 patients. Both drugs were equally effective and well tolerated, but fluvoxamine produced fewer anticholinergic side-effects and caused less sexual dysfunction than clomipramine, but more reports of headache and insomnia.

Paroxetine was of comparable efficacy to clomipramine and both were significantly more effective than placebo in a multinational double-blind placebo-controlled parallel group study of 399 patients with OCD.[71] Bisserbe *et al.*[72] reported that sertraline (50–200 mg/day) was significantly more effective than clomipramine (50–200 mg/day) in a double-blind study ($n = 160$).

Other pharmacological approaches and neurosurgery

Considered as one of the anxiety disorders according to DSM-IV (but not according to the ICD-10), it is not surprising that anxiolytics have been suggested for the treatment of patients with OCD. Thus, clonazepam has been reported as efficient in several uncontrolled studies and case series, and even in a small double-blind randomized multiple crossover study. However, since OCD is a chronic disorder, the long-term use of anxiolytics raises questions of dependency.

Despite reports in open studies regarding the efficacy of trazodone, buspirone, and lithium, results from double-blind studies proved negative. Adding drugs that affect dopamine function, especially the atypical antipsychotics (i.e. risperidone or quetiapine), to SRI therapy in patients with treatment-resistant OCD, may result in improvement for patients with a personal or family history of tics.[53] A combination of SSRI and small doses of high potent dopamine blocker (haloperidol or pimozide) was found to be useful for both the tics and the OC symptoms.[73]

Neurosurgery and Deep Brain Stimulation (DBS) have been reported to be effective in some patients with OCD. Neurosurgery involves procedures that disconnect the outflow pathways originating from the orbitofrontal cortex. Cingulotomy can help some intractable patients, but although the immediate results may be striking, the long-term prognosis is more reserved.[74] As for DBS, initial reports are optimistic[75] but as the total number of patients who underwent this procedure is very small, its efficacy needs to be further elucidated.

Summary of drug treatment for OCD

The first-line treatment consists of either an SSRI or clomipramine. Any one of the six SSRIs (fluoxetine, fluvoxamine, paroxetine, sertraline, citalopram, escitalopram) in current use constitutes an effective and safe choice, but choosing which SSRI depends on the

drug's pharmacokinetic profile, as well as the physician's familiarity with the drug. The dose should be higher than that used for treating depression (e.g. 40–60 mg of fluoxetine) and the trial should last at least 12 weeks. If clomipramine is chosen, cardiovascular problems and closed-angle glaucoma should first be ruled out. Doses of 200 to 300 mg of clomipramine are needed, but titration should last for 1 to 3 weeks and the optimal dose should be continued for at least 12 weeks before determining a lack of response.

If the patient cannot tolerate the first drug (an SSRI) or did not respond, a trial of another SSRI or CMI is advised or augmentation (i.e. risperidone) is recommended.

The third stage in non-responders, and in cases of only a partial response, includes small doses of antipsychotics (especially in Tourette's syndrome), or the addition of lithium or trazodone, buspirone, or tryptophan. The fourth stage consists of atypical neuroleptics, thyroid supplementation, clonidine, a monoamine oxidase inhibitor, and intravenous clomipramine. In truly severe and resistant cases, neurosurgery or DBS could be tried.

Psychological approaches

The effect of a psychodynamic approach in OCD is limited, whereas modern interventions like cognitive and behavioural therapy show promising results.[76] Behavioural therapy was found to be effective in OCD,[76,77] and some data indicate that the beneficial effects of behavioural therapy may be longer lasting.[78] About two-thirds of patients with moderately severe rituals can be expected to improve substantially, but not completely. A combination of behavioural therapy and pharmacotherapy may constitute the optimal treatment for OCD. Recently, two neuroimaging studies found that patients with OCD who are successfully treated with behavioural therapy show changes in cerebral metabolism similar to those produced by successful treatment with SRIs.[47,79]

Behavioural therapy can be conducted in in- and outpatient settings. The principal behavioural approaches in OCD are exposure for obsessions and response prevention for rituals (see Chapter 6.3.2.1). Desensitization, thought stopping, flooding, implosion therapy, and aversion conditioning have also been used in patients with OCD. In behavioural therapy the patient must collaborate and perform assignments. In a study of 18 patients with OCD, those who received exposure and response prevention therapy showed significant improvement, whereas patients on a general anxiety management intervention (control) showed no improvement from baseline.[80] Direct comparisons of behavioural therapy and pharmacotherapy are few and are limited by methodological issues.

In thought stopping, the patient (or initially the therapist) shouts 'stop' or applies an aversive stimulus to counteract the obsessional preoccupation. The patient may also imagine a stop sign with a police officer nearby or another image that evokes inhibition at the same time that he or she recognizes the presence of the obsession. Another technique is to 'postpone' the thought until a specified time (e.g. an hour later) and not to think about it until then.

Psychological factors might be of considerable benefit in understanding what precipitates exacerbations of the disorder and in treating various forms of resistances to treatment, such as non-compliance to medications or to homework assignments. It is important to remember that the symptoms may have important psychological meanings that make patients reluctant to give them up.

In the absence of controlled studies of insight-oriented psychotherapy for OCD, the anecdotal reports reporting lasting change do not allow generalizations to be made regarding efficacy. Also, the efficacy of medications in producing quick improvement has rendered slow and long-term psychotherapy out of favour.

Supportive psychotherapy has a non-specific place in managing patients with OCD, and may help patients improve their functioning and adjustment. The management plan should also include attention to the family members through the provision of emotional support, reassurance, explanation, and advice on how to manage and respond to the patient. Family therapy may reduce marital discord and build a treatment alliance, as well as helping in the resistance to compulsions. Group therapy is useful as a support system for some patients.

Summary

The treatment of OCD was characterized by pessimism until 25 years ago when effective treatments including behavioural therapy and the serotonin reuptake inhibitors were developed. Although introduced for OCD in 1967, it was only in the 1980s that double-blind studies confirmed the efficacy of clomipramine, an SRI. This was followed by the introduction of the selective serotonin reuptake inhibitors, which also proved effective for OCD. The anti-obsessive activity of these drugs was found to be independent from the drugs' antidepressant effect, as established by their efficacy both in depressed and non-depressed patients. Overall, serotonergic therapies have provided a better outlook for these patients and have contributed to our understanding of the pathophysiology of OCD.[6,7] Previously thought to be a rare and untreatable disorder, OCD is now recognized as common, and there is good reason to expect that patients with OCD will benefit substantially from potent SRIs and behaviour therapy.

Many patients with OCD do not seek treatment and the disease tends to be chronic. There is about a 10-year lag between the onset of symptoms and the seeking of professional help due to feelings of embarrassment. Further delay ensues until the diagnosis and correct treatment are given.[81] Census data suggest that over $8 billion are spent in the United States each year on the management of OCD, one-fifth of that spent on cardiac disease.[82] Because patients with OCD often attempt to conceal their symptoms, it is incumbent on clinicians to screen for OCD in every mental status examination, since appropriate treatment can result in improved quality of life, reduced OCD chronicity, and a decrease in cost to the individual and society.

Further information

Kasper, S., Zohar, J., and Stein, D.J. (2002). *Decision making in psychopharmacology*. Martin Dunitz, London.

Montgomery, S. and Zohar, J. (1999). *Obsessive compulsive disorder*. Martin Dunitz, London.

Zohar, J., Hollander, E., Stein, D.J., and Westenberg, H.G. (2007). The Cape Town consensus group. From obsessive-compulsive spectrum to obsessive-compulsive disorders: the Cape Town consensus statement. *CNS Spectrums*, **12**(2 Suppl. 3).

http://www.ocfoundation.org

References

1. Salzman, L. and Thaler, F.H. (1981). Obsessive compulsive disorder: a review of the literature. *The American Journal of Psychiatry*, **138**, 286–96.

2. Robins, L.N., Helzer, J.E., Weissman, M.M., *et al.* (1984). Lifetime prevalence of specific psychiatric disorders in three sites. *Archives of General Psychiatry*, **41**, 949–58.
3. Weissman, M.M., Bland, R.C., Canino, G.J., *et al.* (1994). The cross national epidemiology of obsessive compulsive disorder. *The Journal of Clinical Psychiatry*, **55**, (Suppl. 3), 5–10.
4. Renynghe de Voxrie, G.V. (1968). Anafranil (G34586) in obsessive compulsive neurosis. *Archives Belges de Neurologie*, **68**, 787–92.
5. Fernandez-Cordoba, E. and Lopez-Ibor, A.J. (1967). La monoclorimipramina en enfermos psiquiatricos resistenses a otros tratamientos. *Actas Lusoespañolas de Neurologia, Psiquiatria y Ciencias Afines*, **26**, 119–47.
6. Dolberg, O.T., Iancu, I., Sasson, Y., *et al.* (1996). The pathogenesis and treatment of obsessive-compulsive disorder. *Clinical Neuropharmacology*, **19**, 129–47.
7. Zohar, J. and Insel, T. (1987). Obsessive-compulsive disorder: psychobiological approaches to diagnosis, treatment, and pathophysiology. *Biological Psychiatry*, **22**, 667–87.
8. Goodman, W.K., Price, L.H., Delgado, P.L., *et al.* (1990). Specificity of serotonin reuptake inhibitors in the treatment of obsessive compulsive disorder: comparison of fluvoxamine and desipramine. *Archives of General Psychiatry*, **47**, 577–85.
9. Leonard, H., Swedo, S.E., Lenane, M.C., *et al.* (1991). A double-blind desipramine substitution during long-term clomipramine treatment in children and adolescents with obsessive-compulsive disorder. *Archives of General Psychiatry*, **48**, 922–7.
10. Karno, M., Golding, J.M., Sorenson, S.B., *et al.* (1988). The epidemiology of obsessive-compulsive disorder in five US communities. *Archives of General Psychiatry*, **45**, 1094–9.
11. Flament, M.F., Whitaker, A., Rapoport, J.L., *et al.* (1988). Obsessive compulsive disorder in adolescence: an epidemiological study. *Journal of the American Academy of Child and Adolescent Psychiatry*, **27**, 764–71.
12. Nelson, E. and Rice, J. (1997). Stability of diagnosis of obsessive compulsive disorder in the epidemiologic catchment area study. *The American Journal of Psychiatry*, **154**, 826–31.
13. Stein, M.B., Forde, D.R., Anderson, G., *et al.* (1997). Obsessive compulsive disorder in the community: an epidemiologic survey with clinical reappraisal. *The American Journal of Psychiatry*, **154**, 1120–6.
14. Rasmussen, S.A. and Eisen, J.L. (1992). Epidemiology and clinical features of obsessive-compulsive disorder. In *Obsessive compulsive disorders. Theory and management* (eds. M.A. Jenike, L. Baer, and W.E. Minichiello), pp. 10–27. Year Book, Chicago, IL.
15. Sasson, Y., Dekel, S., Chopra, M., *et al.* (2005). Posttraumatic obsessive compulsive disorder—a case series. *Psychiatry Research*, **135**, 145–52.
16. Pauls, D. (1992). The genetics of OCD and Gilles de la Tourette's syndrome. *The Psychiatric Clinics of North America*, **15**, 759–66.
17. Mataix-Cols, D., Rosario-Campos, M.C., and Leckman, J.F. (2005). A multidimensional model of obsessive-compulsive disorder. *The American Journal of Psychiatry*, **162**, 228–38.
18. Leckman, J.F., Grice, D.E., Boardman, J., *et al.* (1997). Symptoms of obsessive-compulsive disorder. *The American Journal of Psychiatry*, **154**, 911–17.
19. Berman, I., Sapers, B.L., Chang, H.H.J., *et al.* (1995). Treatment of obsessive-compulsive symptoms in schizophrenic patients with clomipramine. *Journal of Clinical Psychopharmacology*, **15**, 206–10.
20. Zohar, J. (1997). Is there room for a new diagnostic subtype—the schizo-obsessive subtype? *CNS Spectrums*, **2**, 49–50.
21. Ravizza, L., Maina, G., and Bogetto, F. (1997). Episodic and chronic OCD. *Depression and Anxiety*, **6**, 154–8.
22. Mataix-Cols, D., Wooderson, S., Lawrence, N., *et al.* (2004). Distinct neural correlates of washing, checking, and hoarding symptom dimensions in obsessive-compulsive disorder. *Archives of General Psychiatry*, **61**, 564–76.
23. Mataix-Cols, D. (2006). Deconstructing obsessive-compulsive disorder: a multidimensional perspective. *Current Opinion in Psychiatry*, **19**, 84–9.
24. Iraqi, Z., El Yazaji, M., Hjiej, H., *et al.* (2005). Obsessive and compulsive symptoms in schizophrenia. Presented in the XIII WPA congress, Cairo-Egypt.
25. Thoren, P., Asberg, M., Gronholm, B., *et al.* (1980). Clomipramine treatment of obsessive compulsive disorder. II. Biochemical aspects. *Archives of General Psychiatry*, **27**, 1289–94.
26. Insel, T.R., Mueller, E.A., Alterman, I., *et al.* (1985). Obsessive compulsive disorder and serotonin: is there a connection? *Biological Psychiatry*, **20**, 1174–88.
27. Weizman, A., Carmi, M., Hermesh, H., *et al.* (1986). High affinity imipramine binding and serotonin uptake in platelets of eight adolescent and ten adult obsessive-compulsive patients. *The American Journal of Psychiatry*, **143**, 335–9.
28. Marazziti, D., Hollander, E., Lensi, P., *et al.* (1992). Peripheral markers of serotonin and dopamine function in obsessive-compulsive disorder. *Psychiatry Research*, **42**, 41–51.
29. Vitiello, B., Shimon, H., Behar, D., *et al.* (1991). Platelet imipramine binding and serotonin uptake in obsessive-compulsive patients. *Acta Psychiatrica Scandinavica*, **84**, 29–32.
30. Kim, S.W., Dysken, M.W., Pandey, G.N., *et al.* (1991). Platelet 3H-imipramine binding sites in obsessive compulsive behavior. *Biological Psychiatry*, **30**, 467–74.
31. Marazziti, D., Pfanner, C., Palego, L., *et al.* (1997). Changes in platelet markers of obsessive compulsive patients during a double-blind trial of fluvoxamine versus clomipramine. *Pharmacopsychiatry*, **30**, 245–9.
32. Flament, M.F., Rapoport, J.L., Murphy, D.L., *et al.* (1987). Biochemical changes during clomipramine treatment of childhood obsessive-compulsive disorder. *Archives of General Psychiatry*, **44**, 219–25.
33. Charney, D.S., Goodman, W.K., Price, L.H., *et al.* (1988). Serotonin function in obsessive-compulsive disorder: a comparison of the effects of tryptophan and *m*-chlorophenylpiperazine in patients and healthy subjects. *Archives of General Psychiatry*, **45**, 177–85.
34. Hollander, E., DeCaria, C.M., Nitescu, A., *et al.* (1992). Serotonergic function in obsessive-compulsive disorder: behavioral and neuroendocrine responses to oral *m*-chlorophenylpiperazine and fenfluramine in patients and healthy volunteers. *Archives of General Psychiatry*, **49**, 21–8.
35. Lesch, K.P., Hoh, A., Disselkamp-Tietze, J., *et al.* (1991). 5-Hydroxytryptamine 1A receptor responsivity in obsessive compulsive disorder. Comparison of patients and controls. *Archives of General Psychiatry*, **48**, 540–7.
36. Bastani, B., Nash, J.F., and Meltzer, H.Y. (1990). Prolactin and cortisol responses to MK-212, a serotonin agonist, in obsessive-compulsive disorder. *Archives of General Psychiatry*, **47**, 833–9.
37. Sasson, Y. and Zohar, J. (1996). New developments in obsessive-compulsive disorder research: implications for clinical management. *International Clinical Psychopharmacology*, **11**(Suppl. 5), 3–12.
38. Benkelfat, C., Murphy, D.L., Zohar, J., *et al.* (1989). Clomipramine in obsessive compulsive disorder: further evidence for a serotonergic mechanism of action. *Archives of General Psychiatry*, **46**, 23–8.
39. Denys, D., Zohar, J., and Westenberg, H.G. (2004). The role of dopamine in obsessive-compulsive disorder: preclinical and clinical evidence. *The Journal of Clinical Psychiatry*, **65**(Suppl. 14), 11–7.
40. Marazziti, D., Dell'Osso, L., Presta, S., *et al.* (1999). Platelet [3H]paroxetine binding in patients with OCD-related disorders. *Psychiatry Research*, **89**, 223–8.

41. McDougle, C.J., Goodman, W.K., Leckman, J.F., *et al.* (1994). Haloperidol addition in fluvoxamine-refractory obsessive-compulsive disorder. A double-blind, placebo-controlled study in patients with and without tics. *Archives of General Psychiatry*, **51**, 302–8.
42. Swedo, S.E., Leonard, H.L., and Kiessling, L.S. (1994). Speculations on antineuronal antibody-mediated neuropsychiatric disorders of childhood. *Pediatrics*, **93**, 323–6.
43. Roy, B.F., Benkelphat, C., Hill, J.L., *et al.* (1994). Serum antibody for somatostatin, 14 and prodynorphin 209–240 in patients with obsessive-compulsive disorder, schizophrenia, Alzheimer's disease, multiple sclerosis and advanced HIV infection. *Biological Psychiatry*, **35**, 335–44.
44. Swedo, S.E., Leonard, H.L., Mittelman, B.B., *et al.* (1997). Identification of children with pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections by a marker associated with rheumatic fever. *The American Journal of Psychiatry*, **154**, 110–12.
45. Rauch, S.L. (1998). Neuroimaging in OCD: clinical implications. *CNS Spectrum*, **3**, (Suppl. 1), 26–9.
46. Baxter, L.R. Jr, Schwartz, J.M., Bergman, K.S., *et al.* (1992). Caudate glucose metabolic rate changes with both drug and behaviour therapy for OCD. *Archives of General Psychiatry*, **49**, 681–9.
47. Galderisi, S., Mucci, A., and Catapano, F. (1995). Neuropsychological slowness in obsessive-compulsive patients: is it confined to tests involving the fronto-subcortical systems? *The British Journal of Psychiatry*, **167**, 394–8.
48. Towey, J.P., Tenke, C.E., Bruder, G.E., *et al.* (1994). Brain event-related potential correlates of over focused attention in obsessive-compulsive disorder. *Psychophysiology*, **31**, 535–43.
49. Hendler, T., Lustig, M., Goshen, E., *et al.* (2003). Brain reactivity to specific symptom provocation indicates prospective therapeutic outcome in OCD. *Psychiatry research*, **124**, 87–103.
50. Rasmussen, S.A. and Tsuang, M.T. (1986). Clinical characteristics and family history in DSM-III obsessive-compulsive disorder. *The American Journal of Psychiatry*, **143**, 317–22.
51. Lenane, M.C., Swedo, S.E., Leonard, H., *et al.* (1990). Psychiatric disorders in first degree relatives of children and adolescents with obsessive-compulsive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **29**, 407–12.
52. Billet, E.A., Richter, M.A., King, N., *et al.* (1997). Obsessive compulsive disorder, response to serotonin reuptake inhibitors and the serotonin transporter gene. *Molecular Psychiatry*, **2**, 403–6.
53. Mundo, E., Richter, M.A., Zai, G., *et al.* (2002). 5HT1Dbeta receptor gene implicated in the pathogenesis of obsessive-compulsive disorder: further evidence from a family-based association study. *Molecular Psychiatry*, **7**, 805–9.
54. Camarena, B., Aguilar, A., Loyzaga, C., *et al.* (2004). A family-based association study of the 5-HT-1Dbeta receptor gene in obsessive-compulsive disorder. *The International Journal of Neuropsychopharmacology*, **7**, 49–53.
55. Insel, T.R., Gillin, J.C., Moore, A., *et al.* (1982). The sleep of patients with OCD. *Archives of General Psychiatry*, **39**, 1372–7.
56. Insel, T.R., Kalin, N.H., Guttmacher, L.B., *et al.* (1982). The dexamethasone suppression test in patients with primary OCD. *Psychiatry Research*, **6**, 153–8.
57. Rasmussen, S.A., Eisen, J.L., and Pato, M.T. (1993). Current issues in the pharmacological management of obsessive-compulsive disorder. *The Journal of Clinical Psychiatry*, **54**, 4s–9s.
58. Clomipramine Collaborative Study Group. (1991). Clomipramine in the treatment of patients with obsessive-compulsive disorder. *Archives of General Psychiatry*, **48**, 730–8.
59. Geller, D.A., Biederman, J., Stewart, S.E., *et al.* (2003). Which SSRI? A meta-analysis of pharmacotherapy trials in pediatric obsessive-compulsive disorder. *The American Journal of Psychiatry*, **160**, 1919–28.
60. Greist, J.H., Bandelow, B., Hollander, E., *et al.* (2003). World council of anxiety. WCA recommendations for the long-term treatment of obsessive-compulsive disorder in adults. *CNS Spectrums*, **8**, 7–16.
61. Kasper, S., Zohar, J., and Stein, D.J. (2002). *Decision making in psychopharmacology*. Martin Dunitz, London.
62. Blier, P., Habib, R., and Flament, M.F. (2006). Pharmacotherapies in the management of obsessive-compulsive disorder. *Canadian Journal of Psychiatry*, **51**, 417–30.
63. Pato, M.T., Zohar-Kadouch, R., Zohar, J., *et al.* (1988). Return of symptoms after discontinuation of clomipramine in patients with obsessive compulsive disorder. *The American Journal of Psychiatry*, **145**, 1521–5.
64. Mundo, E., Barregi, S.R., Pirola, R., *et al.* (1997). Long-term pharmacotherapy of obsessive-compulsive disorder: a double-blind controlled study. *Journal of Clinical Psychopharmacology*, **17**, 4–10.
65. March, J.S., Frances, A., Carpenter, D., *et al.* (1997). The expert consensus guideline series: treatment of obsessive-compulsive disorder. *The Journal of Clinical Psychiatry*, **58**(Suppl. 4), 1–72
66. Montgomery, S.A., Kasper, S., Stein, D.J., *et al.* (2001). Citalopram 20 mg, 40 mg and 60 mg are all effective and well tolerated compared with placebo in obsessive-compulsive disorder. *International Clinical Psychopharmacology*, **16**, 75–86.
67. El Mansari, M. and Blier, P. (2006). Mechanisms of action of current and potential pharmacotherapies of obsessive-compulsive disorder. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, **30**, 362–73.
68. Blier, P. and Szabo, S.T. (2005). Potential mechanisms of action of atypical antipsychotic medications in treatment-resistant depression and anxiety. *The Journal of Clinical Psychiatry*, **66**(Suppl. 8), 30–40.
69. Pigott, T.A., Pato, M.T., Bernstein, S.E., *et al.* (1990). Controlled comparisons of clomipramine and fluoxetine in the treatment of obsessive-compulsive disorder: behavioral and biological results. *Archives of General Psychiatry*, **47**, 926–32.
70. Freeman, C.P.L., Trimble, M.R., Deakin, J.F.W., *et al.* (1994). Fluvoxamine versus clomipramine in the treatment of obsessive compulsive disorder: a multi-center, randomized, double-blind, parallel group comparison. *The Journal of Clinical Psychiatry*, **55**, 301–5.
71. Zohar, J. and Judge, R. (1996). Paroxetine versus clomipramine in the treatment of obsessive-compulsive disorder. *The British Journal of Psychiatry*, **169**, 468–74.
72. Bisserbe, J.C., Lane, R.M., Flament, M.F., *et al.* (1997). A double-blind comparison of sertraline and clomipramine in outpatients with obsessive-compulsive disorder. *European Psychiatry*, **153**, 1450–4.
73. McDougle, J., Goodman, W.K., Price, L.H., *et al.* (1990). Neuroleptic addition in fluvoxamine-refractory obsessive-compulsive disorder. *The American Journal of Psychiatry*, **147**, 652–4.
74. Jenike, M.A., Baer, L., Ballantine, T., *et al.* (1991). Cingulotomy for refractory obsessive-compulsive disorder. *Archives of General Psychiatry*, **48**, 548–55.
75. Greenberg, B.D., Malone, D.A., Friehs, G.M., *et al.* (2006). Three-year outcomes in deep brain stimulation for highly resistant obsessive-compulsive disorder. *Neuropsychopharmacology*. **31**, 2384–93.
76. Van Balkom, A.J.L.M., De Haan, E., Van Oppen, P., *et al.* (1998). Cognitive and behavioral therapies alone versus in combination with fluvoxamine in the treatment of obsessive compulsive disorder. *The Journal of Nervous and Mental Disease*, **186**, 492–9.
77. Marks, I.M., Hodgson, R., Rachman, S., *et al.* (1975). Treatment of chronic obsessive-compulsive neurosis *in vivo* exposure: a 2-year follow-up and issues in treatment. *The British Journal of Psychiatry*, **127**, 349–64.

78. Greist, J.H. (1996). New developments in behaviour therapy for obsessive-compulsive disorder. *International Clinical Psychopharmacology*, **11**(Suppl.), 63–73.
79. Schwartz, J.M., Stoessel, P.W., Baxter, L.R. Jr., *et al.* (1996). Systematic changes in cerebral glucose metabolic rate after successful behavior modification treatment of OCD. *Archives of General Psychiatry*, **53**, 109–13.
80. Lindsay, M., Craig, R., and Andrews, G. (1997). Controlled trial of exposure and response prevention in obsessive-compulsive disorder. *The British Journal of Psychiatry*, **171**, 135–9.
81. Hollander, E. (1997). Obsessive compulsive disorder: the hidden epidemic. *The Journal of Clinical Psychiatry*, **12**(Suppl.), 3–6.
82. DuPont, R.L., Rice, D.P., Shiraki, S., *et al.* (1995). Economic costs of obsessive compulsive disorder. *Medical Interface*, **8**, 102–9.

4.9

Depersonalization disorder

Nick Medford, Mauricio Sierra,
and Anthony S. David

Introduction

Depersonalization, a term coined by Dugas in 1898,[1] is defined in DSM-IV as 'an alteration in the experience of self so that one feels detached from and as if one is an outside observer of one's outside mental processes or body'. Brief, self-limiting experiences of depersonalization commonly occur in healthy people in the context of fatigue, intense stress, or during/after intoxication with alcohol or illicit drugs. However, some people experience chronic depersonalization of a disturbing intensity, causing significant distress and impacting on quality-of-life and daily functioning. This may occur as a **primary depersonalization disorder (DPD)**, or in the context of other psychiatric or neurological conditions. In this chapter, we consider the primary disorder, although some sections are also relevant to secondary depersonalization.

The depersonalization experience is one of feeling strangely altered and unreal, in a way that sufferers often find very hard to convey. It is often accompanied by the related phenomenon of **derealization**, in which the person's surroundings are experienced as somehow remote and lacking immediacy and vibrancy, as if the world itself has become oddly unreal. Patients with persistent depersonalization and derealization often use the analogy of feeling as if they are on the set of a play or film, where nothing is real and they are acting out a role rather than living a real life.

Clinical features

The diagnosis requires the presence of persistent, distressing depersonalization and/or derealization, occurring in clear consciousness, and not due to another disorder or substance. Some patients find it impossible to divide their symptoms into depersonalization and derealization, seeing them as essentially two ways of describing the same experience. Nevertheless, one may encounter patients who describe one without the other. 'Pure' derealization is, however, uncommon.

In addition, there are a number of other symptoms that occur with sufficient frequency to be considered as part of the depersonalization syndrome, although their presence is not essential for making the diagnosis. These are as follows:

Desomatization—a loss or diminution of bodily sensation, sometimes accompanied by a feeling of disembodiment.

De-affectualization—a loss or diminution of emotional reactivity-the feeling that life has somehow been drained of emotional content, or that the sufferer feels little emotion in response to people or events that would normally be expected to elicit an emotional response. This may have significance for intimate relationships. It should be noted that de-affectualization is not usually accompanied by blunted affect of the type commonly seen in schizophrenia.

De-ideation—a feeling of mental emptiness which may cause difficulty in concentrating, a distorted experience of time, and a sense of detachment from memories. Often accompanied by a feeling of 'stuffiness in the head' or 'as if my brain has turned to cotton wool'.

While 'depersonalization' and 'derealization' are well-established in the psychiatric lexicon, the three terms listed above are not widely used or discussed. However, a recent analysis of symptoms reported by patients with DPD gave strong support to the idea that the condition should be considered as a syndrome, with symptoms occurring in domains corresponding to the terms used above.[2]

Classification

In DSM-IV, DPD is classified as a dissociative disorder, while in ICD-10 it falls under the vague heading of 'other neurotic disorders', and is not linked to any other category of disorder.

It has been argued that DPD is not truly a dissociative disorder, as dissociation is generally characterized by a lack of subjective awareness of change, whereas in DPD the experience of feeling changed is central. However, this apparent contradiction can be resolved if dissociation is conceptualized as a category incorporating both types of phenomenon.[3]

Diagnosis and differential diagnosis

The diagnosis should be established by a careful clinical assessment. Because DPD remains a somewhat obscure disorder, patients with this condition may have had unproductive consultations with other professionals and formed the impression that their symptoms are baffling, perhaps even unique. Being given the correct diagnosis and the opportunity to discuss it in depth with an informed psychiatrist may come as a great relief. The reassurance derived from this may in itself have a powerful therapeutic effect.

Where there are other psychiatric symptoms (e.g. anxiety, panic attacks, depressive features), the distinction between primary and secondary depersonalization may be difficult. The best way to approach this is simply to establish what the dominant symptoms are at the time of presentation. In a patient with a history of panic disorder who has developed severe unremitting depersonalization, and now has very infrequent panic attacks, the most pragmatic approach is to diagnose DPD. The fact that the panic symptoms preceded the onset of DPD is less important than the fact that it is now the DPD symptoms that dominate the clinical picture.

This issue aside, the main psychiatric differential hinges on the possibility that when patients describe feeling altered or unreal, they are articulating delusional beliefs. It is important to establish that patients have no psychotic symptoms; in particular, that they do not literally believe themselves to be unreal or dead, as this is suggestive of psychotic depression and the Cotard delusion, rather than DPD.

Depersonalization can also occur in neurological disorders, principally temporal lobe epilepsy and migraine.[4] Here the history is usually of brief, stereotyped episodes, with associated features that should provide sufficient clues to the underlying diagnosis.

Epidemiology

Until recently, DPD was considered rare, but contemporary epidemiological work suggests that it affects 1–2 per cent of the general adult population,[5] with a gender ratio of 1:1.

Symptom surveys suggest that depersonalization is perhaps the third commonest symptom (after anxiety and low mood) in psychiatric populations. It should be noted that these studies do not distinguish between primary and secondary depersonalization.

Aetiology

Various factors have been implicated in the genesis and maintenance of the condition. Biological and psychological issues are considered separately here, but should not be seen as mutually exclusive.

Psychological factors: Many patients with primary DPD have concurrent anxiety or mood symptoms, or a previous history of anxiety and/or panic attacks. The clinical impression is often that feelings of detachment and unreality have arisen as a defence against feeling anxious and threatened—a way of keeping a stressful world at a safer psychological distance. There is sometimes a history of DPD symptoms first occurring in the context of some particularly stressful event or period. Often, however, specific precipitants are not identifiable.

Once DPD develops, further anxiety may follow—patients may worry that, for example, the peculiar feeling of unreality is a sign that they are on the verge of mental breakdown. These concerns often manifest as obsessional rumination and self-monitoring, characterized by a compulsive checking of the inner state and a comparison of this state with some idealized standard of normality. This further anxiety can lead to reinforcement and perpetuation of depersonalization and derealization, so that symptoms feed each other in a 'vicious circle'.

Biological factors: There is objective biological evidence relating to the loss of emotional reactivity that patients with DPD commonly report. Patients with DPD show attenuated skin conductance responses to emotional stimuli,[6] while fMRI work suggests that the brain's emotional response circuitry is inhibited in DPD.[7,8]

Some 10–20 per cent of patients with DPD describe symptoms beginning during or after an episode of illicit drug use, cannabis being the drug most commonly implicated. Many illicit drugs are known to cause depersonalization phenomena acutely, but it is unclear how any drug might produce chronic symptoms persisting long after the drug is cleared from the system. It seems likely that the initial symptoms are due to the drug, but become chronic and unremitting through the kind of "vicious circle" outlined above.[9]

Course and prognosis

Two large case series[10,11] suggest that age of onset is usually in late adolescence or early adulthood, although up to a third of patients describe symptoms originating in childhood. Onset may be sudden or gradual, and symptoms thereafter may be episodic or continuous. Patients often report little or no fluctuation in symptom nature or severity, although with close questioning it is often possible to establish that certain factors (e.g. stress, fatigue) worsen the symptoms.

Symptoms are often unremitting for many years. In the largest case series to date,[10] patients had been symptomatic for a mean of 13.9 years at the time of initial presentation to a specialist DPD clinic. This striking statistic may, in part, reflect the widespread lack of familiarity with the condition and consequent delays in diagnosis.

While there is little available data on which to base predictions about prognosis, symptoms that have been continuous for many years tend to be more refractory than those of more recent origin.

Evaluation of treatments

Until recently, the treatment literature consisted of small case series or single case reports, but a few larger studies have now been performed. Key findings are:

Pharmacological treatments

Lamotrigine: Despite encouraging results from a pilot study, a placebo-controlled crossover trial did not show evidence of efficacy. However, in a larger, more recent open-label study of lamotrigine-antidepressant combination therapy, significant improvements were seen in a majority of patients.[12]

Opioid antagonists: Transient reductions in symptoms have been reported in response to naloxone infusion, and a recent open-label study of oral naltrexone showed some evidence of efficacy.[13]

Clomipramine: One small open-label study, results inconclusive.[14]

There remains a paucity of data from rigorous controlled trials. To date, the only large double-blind randomised controlled trial is a study of *fluoxetine*, which found no evidence of efficacy.[15]

Psychological treatments

Despite a number of reports of successful treatment with a range of psychotherapeutic techniques, the only treatment trial is an open study of cognitive behavioural therapy (CBT), which showed significant clinical benefits.[16]

Management

It will be appreciated from the above that there is insufficient evidence on which to base definitive treatment guidelines, and as yet

no drugs are licensed for the treatment of DPD in the UK. Current treatment strategies are based on encouraging results from exploratory studies, rather than on any overwhelming weight of evidence, and this limitation should not be concealed from patients.

The combination of lamotrigine (up to 500 mg per day) and an SSRI (usually citalopram or escitalopram) is often used as first-line treatment (see Medford *et al.* in Further Reading below). An alternative is clonazepam (0.5–4 mg per day).There have been no clinical trials of clonazepam in DPD, but many patients find it helpful, particularly when there is co-morbid anxiety. The risk of dependency must be carefully weighed against possible clinical benefits.

Naltrexone (see above) may also have a role, while clomipramine may be helpful when obsessional ruminations are prominent. There is anecdotal evidence to support the use of bupropion in treatment-refractory cases.

CBT may be beneficial, either alone or in combination with pharmacotherapy. Depersonalization experiences do not readily lend themselves to a cognitive behavioural analysis, but CBT techniques may be helpful in addressing associated anxieties, ruminations, and avoidance behaviours, and use of CBT should be considered whenever any of these features are prominent.

The use of standardized rating scales can assist in diagnosis and monitoring response to treatment. The Cambridge Depersonalization Scale[17] is particularly recommended.

In secondary depersonalization, it is usually appropriate to simply pursue conventional treatment of the primary condition, but if depersonalization becomes severe and disabling it may be necessary to treat it more specifically. Treatment of depersonalization in the context of substance misuse is particularly difficult: since most drugs of abuse can cause or exacerbate the symptoms, there is generally little point in attempting specific treatment unless abstinence has been established.

Further information

Mayer-Gross, W. (1935). On depersonalization. *British Journal of Medical Psychology*, **15**, 103–22. Classic early monograph, still of relevance.

Medford, N., Sierra, M., Baker, D., *et al.* (2005). Understanding and treating depersonalisation disorder. *Advances in Psychiatric Treatment*, **11**, 92–100. Conceptual and clinical overview by clinicians from the UK's only specialized DPD clinic.

Schilder, P. (1950). The Image and Appearance of the Human Body. New York: International Universities Press. Contains a fascinating psychodynamic account of DPD.

Simeon, D. andAbugel, J. (2006). Feeling Unreal: Depersonalization Disorder and the Loss of the Self. Oxford University Press, USA. Readable and informative book on the disorder.

References

1. Sierra, M. and Berrios, G.E. (1996) Un cas de depersonnalisation, by L.Dugas. Translation and introduction. *History of Psychiatry*, **7**, 451–61.
2. Sierra, M., Baker, D., Medford, N., *et al.* (2005). Unpacking the depersonalization syndrome: an exploratory factor analysis on the Cambridge Depersonalization Scale. *Psychological Medicine*, **35**, 1523–32.
3. Holmes, E.A., Brown, R.J., Mansell, W., *et al.* (2005). Are there two qualitatively distinct forms of dissociation? A review and some clinical implications. *Clinical Psychology Review*, **25**, 1–23.
4. Lambert, M.V., Sierra, M., Phillips, M.L., *et al.* (2002). The spectrum of organic depersonalisation: a review plus four new cases. *Journal of Neuropsychiatry and Clinical Neurosciences*, **14**, 141–54.
5. Hunter, E.C.M., Sierra, M. and David, A.S. (2004). The epidemiology of depersonalization and derealisation: a systematic review. *Social Psychiatry Psychiatric Epidemiology*, **39**, 9–18.
6. Sierra, M., Senior, C., Dalton, J., *et al.* (2002). Autonomic response in depersonalization disorder. *Archives of General Psychiatry*, **59**, 833–38.
7. Phillips, M., Medford, N., Senior, C., *et al.* (2001). Depersonalization disorder: thinking without feeling. *Psychiatry Res Neuroimaging*, **108**, 145–60.
8. Medford, N., Phillips, M., Brierley, B., *et al.* (2006) Emotional memory in depersonalization disorder. *Psychiatry Res Neuroimaging*, **148**, 93–102.
9. Medford, N., Baker, D., Hunter, E., *et al.* (2003). Depersonalization following illicit drug use: a controlled analysis of 40 cases. *Addiction*, **12**, 1731–6.
10. Baker, D., Hunter, E., Lawrence, E., *et al.* (2003). Depersonalization disorder: clinical features of 204 cases. *British Journal of Psychiatry*, **182**, 428–33.
11. Simeon, D., Knutelska, M., Nelson, D., *et al.* (2003). Feeling unreal: a depersonalization disorder update of 117 cases. *Journal of Clinical Psychology*, **64**, 990–7.
12. Sierra, M., Baker, D., Medford, N., *et al.* (2006). Lamotrigine as an add-on treatment for depersonalization disorder: a retrospective study of 32 cases. *Clinical Neuropharmacology*, **29**, 253–8.
13. Simeon, D. and Knutelska, M. (2005). An open trial of naltrexone in the treatment of depersonalization disorder. *Journal of Clinical Psychopharmacology*, **25**, 267–270.
14. Simeon, D., Stein, D.J., Hollander, E. (1998). Treatment of depersonalization disorder with clomipramine. *Biological Psychiatry*, **44**, 302–03.
15. Simeon, D., Guralnik, O., Schmeidler, J., *et al.* (2004). Fluoxetine therapy in depersonalization disorder: randomized controlled trial. *British Journal of Psychiatry*, **185**, 31–6.
16. Hunter, E., Baker, D., Phillips, M., *et al.* (2005). Cognitive-behaviour therapy for depersonalisation disorder: an open study. *Behaviour Research and Therapy*, **43**, 1121–30.
17. Sierra, M.and Berrios, G.E. (2000). The Cambridge Depersonalization Scale: a new instrument for the measurement of depersonalization. *Psychiatry Research*, **93**, 153–64.

4.10

Disorders of eating

Contents

4.10.1 Anorexia nervosa

Gerald Russell

Introduction: history of ideas

Two different approaches may be discerned in the conceptualization of anorexia nervosa.

1 The medicoclinical approach defines the illness in terms of its clinical manifestations; the main landmarks were the descriptions by William Gull in 1874[1] and Charles Lasègue in 1873.[2]

2 The sociocultural approach is unlike the more empirical clinical approach and takes causation into account by viewing the illness as a response to prevailing social and cultural systems. This was well argued by the social historian Joan Jacobs Brumberg who considers anorexia nervosa simply as a control of appetite in women responding to widely differing forces which may change during historical times.[3]

There is a strong argument for accepting the original descriptions by Gull and Lasègue as containing the essence of anorexia nervosa. They both recognized a disorder associated with severe emaciation and loss of menstrual periods, inexplicable in terms of recognized physical causes of wasting. They were both extremely cautious about the nature of the mental disorder. Gull spoke of a morbid mental state or 'mental perversity', and adopted the more general term anorexia 'nervosa' which has persisted until today. Lasègue also referred to 'mental perversity' but was bold enough to call the condition 'anorexie hystérique, faute de mieux'.

It is probably best to seek a balance between the diagnostic rectitude of the medicoclinical approach and the malleability of anorexia nervosa in different sociocultural settings. Looking back in historical times, it may well be that the self-starvation and asceticism of St Catherine of Siena corresponded to modern anorexia nervosa.[4] In more recent times the preoccupations of the patients have altered so that their disturbed experience with their own body[5] or their 'morbid fear of fatness'[6] has become one of the diagnostic criteria. Yet this concern with body size was not remarked upon by Gull or Lasègue. This is an argument for concluding that at least the psychological content, and perhaps also the form, of anorexia nervosa are changeable in response to historical times and sociocultural influences.

Epidemiology

Screening instruments

The most commonly used screening test in the detection of anorexia nervosa is the Eating Attitudes Test (**EAT**).[7] Doubt has, however, been expressed about the predictive value of the EAT in the very populations where its use was introduced, as only a small percentage of the EAT-screened positive scores will have an actual eating disorder. Thus, the EAT has limited usefulness in surveys for detecting anorexia nervosa unless it is supplemented by detailed clinical assessments. There is also a risk of failing to detect cases of anorexia nervosa as it was found that patients currently receiving active treatment were among the non-respondents, presumably because they wished to conceal their disorder.[8]

Populations surveyed

(a) General population surveys

These are often impracticable when the aim is to detect anorexia nervosa, a relatively uncommon disorder.

(b) Surveys of primary care populations

A useful compromise is that of surveying populations of patients who consult their general practitioners. A Netherlands study was successful because a large population was surveyed (over 150 000) and the general practitioners themselves, after suitable training, were responsible for making the diagnoses.[9,10]

(c) Populations thought to be more at risk

Surveys of ballet and modelling students were conducted because it was thought likely that there would be a high prevalence of anorexia nervosa among them as a result of pressures exerted to sustain a slim figure in keeping with their professional image.[11,12]

Populations of adolescent schoolgirls have also been surveyed as their susceptibility might be raised by virtue of their age, sex, and the frequency of dieting among the school population.[13] The most thorough survey of 15-year-old schoolchildren was that conducted in Göteborg, Sweden.[14]

(d) Surveys based on case registers and hospital records

Data have been obtained on patients referred to inpatient and outpatient psychiatric services, or with the addition of patients who had consulted paediatricians, general medical services, or gynaecologists.[15,16]

Results of epidemiological surveys

(a) Incidence of anorexia nervosa

The studies which counted only hospitalized patients tended to yield low estimates of the annual incidence of anorexia nervosa expressed per 100 000 population (e.g. 0.45 in Sweden[15]). Estimates based on case registers of psychiatric patients similarly yielded fairly low incidence rates (e.g. 0.64 in Monroe County, New York[15]). The incidence found in community-based studies was by far the highest (7.7 in The Netherlands[10] and 8.2 in Rochester, Minnesota[17]), presumably because they included the less severe cases.

(b) Prevalence of anorexia nervosa in vulnerable populations

A high prevalence rate was found among Canadian ballet students (6.5 per cent) and modelling students (7 per cent).[11] A similar survey in an English ballet school also showed a high prevalence of 'possible' cases of anorexia nervosa (7.0 per cent).[12]

Surveys among schoolgirls have shown a fairly wide variation in prevalence rates, ranging from 0 per cent to 1.1 per cent. In the English studies a consistent difference in prevalence rate was found between private schools (1 per cent) and state schools (0–0.2 per cent).[13] This social class distinction was not so definite in the Swedish study where the overall prevalence of 0.84 per cent of schoolgirls, up to and including 15 years of age, represents a high rate for anorexia nervosa.[14]

(c) Age and sex

Epidemiological surveys have confirmed clinical opinion that anorexia nervosa commences most frequently in the young, especially within a few years of puberty. The peak age of onset is 18 years.[13] The illness is less common before puberty, but in a series of patients admitted to a children's hospital the age of onset ranged from 7.75 to 14.33 years.[18] A prevalence of 2.0 per cent in females aged 18 to 25 was found in a community survey in Padua, Italy.[19]

A marked predominance of females over males is usually reported in surveys, for example 92 per cent in North-east Scotland[15] and 90 per cent among the children in Göteborg.[14] On the other hand a community survey in the province of Ontario gave rise to a more balanced female to male ratio (2:1) when cases of partial anorexia nervosa were included.[20]

(d) Social class and socio-economic status

The view has been widely held that anorexia nervosa occurs predominantly in patients with middle-class backgrounds, since Fenwick's classical observation that anorexia nervosa 'is much more common in wealthier classes of society than amongst those who have to procure their bread by daily labour'.[15]

Epidemiological surveys aimed at wider populations leave the question of social class distribution somewhat equivocal. Whereas a high percentage of combined social classes 1 and 2 (Registrar General's categories) were found in clinical studies,[21] a high social class predominance was not found in studies utilizing case registers.[15] On the other hand, the schoolgirl studies mentioned above tended to confirm a high social class predominance.

Has the incidence of anorexia nervosa increased since the 1950s?

From the 1970s experienced clinicians have expressed their view that anorexia nervosa had increased in frequency. This view gained support from surveys repeated on the same population after intervals of 10 years or more in Sweden, Scotland, Switzerland, and Monroe County, New York.[15] Most investigations found a clear trend for an increased incidence over time although it appears that a plateau was reached in the 1980s.[16] In the most thorough study, from Rochester, Minnesota[17] the increase was confined to female patients aged 15 to 24 years. There were similar findings from the Netherlands with a rise from 56.4 to 109.2 per 100 000 among 15- to 19-year-old females from 1985–1989 to 1995–1999.[10] Although there is support for an increased incidence of anorexia nervosa, there remain dissenting voices.

It is better to pose a different question which renders any controversy unnecessary. We should ask instead whether there has been an increase in the incidence of eating disorders including anorexia nervosa. This is especially relevant in view of the fact that bulimia nervosa is a variant of anorexia nervosa.[22] There is strong evidence that bulimia nervosa is a new disorder and has not simply appeared because of improved medical recognition.[23] Moreover, the incidence of bulimia nervosa has risen sharply at least until the mid 1990s, so that it is now about double that of anorexia nervosa[24] (see also Chapter 4.10.2). In conclusion, it is clear that since the 1960s there has been a significant increase in eating disorders, of which the two clearest syndromes are anorexia nervosa and bulimia nervosa.

Aetiology

Aetiological concepts

According to one robust opinion, it is essential to pursue the search for a specific and necessary cause of anorexia nervosa because the currently popular 'multifactorial' approach has little explanatory power. Accordingly the failure to identify a necessary causal element is regrettable. Many of the factors within a wide range of psychological, social, and physical causes so far studied may therefore only be relevant in predisposing to anorexia nervosa, whose causes still elude clarification'.[25]

The multidimensional approach to anorexia nervosa

It is precisely because we do not know the fundamental (necessary) cause of anorexia nervosa that recourse has to be had to a multi-

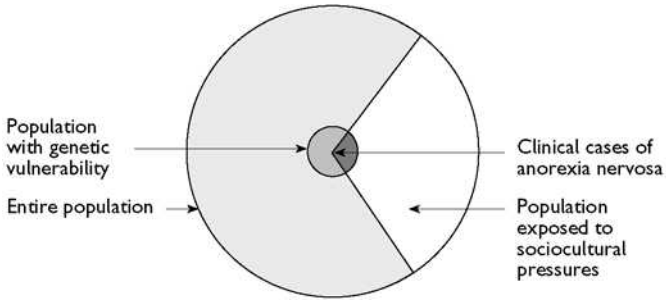


Fig. 4.10.1.1 Diagrammatic illustration of the way that genetic predisposition interacts with sociocultural pressures to cause anorexia nervosa.

dimensional approach, faute de mieux. Although it has its limitations, a multidimensional approach permits one to consider a range of possible causal factors which not only act in an additive manner but may combine in a specific way to bring about the illness: 'It is the interaction and timing of these phenomena in a given individual which are necessary for the person to become ill'.[26]

It is useful to provide a simple model of the way that two broad sets of factors may interact and augment each other (Fig. 4.10.1.1). The outer circle represents the entire population in a developed 'westernized' country. Within the circle there is a large sector representing females within an age range of 10 to 50 years who experience prevailing social pressures to acquire a slender body shape through dieting. Evidently only a small proportion of these women develop the illness. It is likely that for anorexia nervosa to develop it is also necessary to possess a genetic predisposition, represented by the small inner circle. The intersection of the inner circle and the large sector produces a small sector of females who have the genetic predisposition and also experience sociocultural pressures to lose weight, interacting to cause clinical anorexia nervosa.

Causal factors may not only interact as explained above, but they can also influence the content of the illness, its 'colouring', and its form. This modelling function is described by the term 'pathoplastic' which was introduced by Birnbaum.[27] Pathoplastic features are to be distinguished from the more fundamental causes of psychiatric illness, but they do exert a predisposing tendency as well as a modelling role.

Sociocultural causes

(a) The cult of thinness

The pathoplastic sociocultural causes of eating disorders have been subsumed under the term 'the modern cult of thinness' prevalent in westernized societies.[15] Vulnerable patients are likely to respond to this pressure by experimenting with weight-reducing diets which carry a degree of risk, and anorexia nervosa is arguably but an extension of determined dieting.[28]

It is proposed that the cult of thinness is responsible for the increase in the incidence of eating disorders since the 1950s. The social pressures which lead to dietary restraint include the publication of books and magazines advising weight-reducing diets, the fashion industry which caters mainly for the slimmer figure, television attaching sexual allure and professional success to the possession of a svelte figure, and the emphasis on physical fitness and athleticism.[3]

(b) Changes in the psychopathology of anorexia nervosa

At the beginning of this chapter it was pointed out that the psychological content and the form of anorexia nervosa have changed over historical time and in response to sociocultural influences. Whereas Gull and Lasègue spoke only of 'anorexia nervosa' and 'anorexie hystérique' respectively, more recent descriptions of the psychopathology of anorexia nervosa have stressed a disturbed experience of one's own body,[5] a weight 'phobia',[29] or a morbid dread of fatness.[6] It is precisely the modern anorectic's dread of fatness that is most in keeping with today's cult of thinness. It is arguable therefore that modern societal pressures have determined the patients' preoccupations and contributed to their food avoidance. These beliefs are held obstinately and amount to overvalued ideas.

(c) Anorexia nervosa as a culture-bound syndrome

The proposition that anorexia nervosa is a culture-bound syndrome has much support.[30,31] An intense fear of becoming obese is as culture bound as the disorder koro (a fear of shrinkage of the genitals) in Malaysia and South China.

A culture-bound syndrome may be defined as a collection of signs and symptoms which is not to be found universally in human populations, but is restricted to a particular culture or group of cultures. Implicit is the view that culture factors play an important role in the genesis of the symptom cluster . . . [30]

Anorexia nervosa meets these criteria, first because it is limited to westernized or industrialized nations, and secondly because it is clear that the psychosocial pressures on women to become thin constitute a powerful cultural force leading to anorexia nervosa.

In order to allow for exceptions to the rule, when anorexia nervosa occurs in non-westernized countries, the illness may be understood as arising from cultures undergoing rapid cultural change.[31] Anorexia nervosa is thus a 'culture-change syndrome', explaining its increased incidence in Japan and Israel. Anorexia nervosa remains rare in Asia (particularly India), the Middle East (with the exception of Israel), and generally in poorly developed countries.

Young female immigrants who move to a new culture may suffer from an increased prevalence of eating disorders. For example, children with anorexia nervosa were found among Asian families living in Britain. This was linked with an exposure to a conflicting set of sociocultural norms in comparison with their parents and grandparents.[32]

Adverse life events

Anorexia nervosa may be precipitated by adverse life events. Early clinical studies depended solely on the patient's reports of an adverse life event preceding the onset of the illness. These varied widely in severity and included the death of grandparents, a father's remarriage, a severe physical illness, stress or failure of school examinations, or being teased about being overweight.[21]

Recent studies have relied on more objective measurements of adverse life events and comparisons with control groups. In one such study,[33] life events rated included the death of a close relative, a poor relationship with parents, or an unhappy marriage. Fairly severe life events and lasting difficulties were found in the majority of patients with a late onset (after 25 years), whereas they were only found in a minority of patients with an early onset.

Anorexia nervosa can also be precipitated by sexual experiences and conflicts. In a series of 31 adolescent and young women it was

observed that in 14 first sexual intercourse had occurred before the illness.[34] Sexual problems were seen by 13 patients as major precipitants of their illness; 10 of them had experienced intercourse. The authors concluded that a specific aetiological role of sexual factors seemed unlikely, but there might be a direct relationship between the onset of eating difficulties and concurrent sexual problems.

In a series of 15 patients, anorexia nervosa developed during pregnancy or, more often, during the post-natal period when it was accompanied by depression.[35] Risk factors included ambivalence about motherhood, a large weight gain during the pregnancy, physical complications during pregnancy, post-natal depression, and past psychiatric illness.

Childhood sexual abuse

Since it was reported that a high proportion of patients in a treatment programme for anorexia nervosa gave histories of sexual abuse in childhood, it has been supposed that this trauma would be a contributory causal factor.[36] It would be better if this history could be corroborated, but for obvious reasons it is often difficult to do so. Hence this subject raises unusual difficulties in judging the reliability of the data. Child sexual abuse is also discussed in the chapter on bulimia nervosa (Chapter 4.10.2).

In a careful study of childhood sexual experiences reported by women with anorexia nervosa, the authors classified the events according to the seriousness of the sexual act in childhood and concentrated on sexual experiences with someone at least 5 years older.[37] They found surprisingly high rates (about one-third) of adverse sexual experiences in women with eating disorders, and, unlike other investigators, they did not find that the frequency of these reports was less in anorexia nervosa than bulimia nervosa. They concluded that it was plausible that childhood sexual contact with an adult may in some cases contribute to a later eating disorder.

The complexity of this subject has been increased by a study on the relationship between sexual abuse, disordered personality, and eating disorders.[38] The authors found that 30 per cent of patients referred to an eating disorders clinic gave a history of childhood sexual abuse. In addition, they found that 52 per cent of their patients had a personality disorder. A significantly higher proportion of women with a personality disorder had a history of childhood sexual abuse, compared with those without a personality disorder. Surprisingly they still concluded that in patients with eating disorders it was not possible to show a clear causal link between child sexual abuse and personality disorder.

In a review of the subject a number of hypotheses were examined on the relationship between childhood sexual abuse and eating disorders.[39] One hypothesis is that child sexual abuse is more common in bulimia nervosa than in anorexia nervosa. The authors had to concede that the findings remain inconclusive. They also examined the question of whether child sexual abuse is a specific risk factor for eating disorders. They concluded that this was not the case, as the rates of child sexual abuse in eating disorder patients were similar to or less than those in various other psychiatric comparison groups. Finally, they found strong evidence that patients with eating disorders reporting child sexual abuse were more likely to show general psychiatric symptoms including depression, alcohol problems, or suicidal gestures, as well as an association with personality disorders.

Family factors

Two influential groups of family therapists (Minuchin at the Philadelphia Child Guidance Clinic and Selvini Palazzoli in Milan) have devised family models to explain the genesis of anorexia nervosa.

Minuchin *et al.* identified faulty patterns of interaction between members of the anorexic patient's family; they in turn were thought to lead to the child's attempt to solve the family problems by starving herself:[40] 'The sick child plays an important role in the family's pattern of conflict avoidance, and this role is an important source of reinforcement for his symptoms'. Five main characteristics of family interaction were identified as detrimental to the function of the family: enmeshment (a tight web of family relationships with the members appearing to read each other's minds), overprotectiveness, rigidity, involvement of the sick child in parental conflicts, lack of resolution of conflicts.

Selvini Palazzoli[41] also identified abnormal patterns of communication within these families and in addition described abnormal relationships between the family members. She assumed that anorexia nervosa amounted to a logical adjustment to an illogical interpersonal system.

Bruch[42] described girls who developed anorexia nervosa as 'good girls', who previously had a profound desire to please their families to the point of becoming unaware of their own needs. The frequency of broken families in anorexia nervosa is thought to be fairly low. Anorexic families have been found to be closer: the patients more often perceived themselves as having happy relationships within the family.

It remains uncertain whether these abnormal interactions are to blame for the illness or develop as a response by parents faced with a starving child. Careful therapists take pains to reassure parents at the commencement of family therapy: ' . . . we always find it useful to spend some time discussing the nature of the illness, stressing in particular that we do not see the family as the origin of the problem'.[43]

Personality disorders

A sizeable proportion of patients (30 per cent[15] and 32 per cent[21]) were said to have had a 'normal' personality during childhood before their illness. Nevertheless there is general agreement of a close relationship between obsessional personalities and the later development of anorexia nervosa. In fact Janet, who carefully described obsessions and psychasthenia, was dubious about the validity of the diagnostic concept of anorexia nervosa. He thought that the patient's fear of fatness was an elaborate obsessional idea.[44]

In a study of patients admitted for treatment they were classified into anorexia nervosa, bulimia nervosa, or a combination of the two disorders.[45] Personality disorders were identified through the Structured Clinical Interview for DSM-IIIR personality disorders (SCID-II). Seventy-two per cent of the patients met the criteria for at least one personality disorder. Anorectics were found to have a high rate of obsessive–compulsive personality disorder.

There have been attempts to disentangle the features of premorbid personality and illness in order to clarify the personality characteristics predisposing to anorexia nervosa. Women who had recovered from restricting anorexia nervosa were tested at an 8- to 10-year follow-up, using a number of self-report instruments.[46] They were compared with two control groups: normal women and the sisters of the recovered anorexic patients. The women who had

recovered rated higher on risk avoidance and conforming to authority. They also showed a greater degree of self-control and impulse control, and less enterprise and spontaneity.

Biomedical factors and pathogenesis

(a) Historical notes

Since the early part of the twentieth century a recurring theme has been the possibility that anorexia nervosa is primarily caused by an endocrine or cerebral disturbance. From 1916 there was much preoccupation with the concept of Simmonds' cachexia,[47] the assumed result of latent disease of the pituitary gland. There was diagnostic confusion between anorexia nervosa and hypopituitarism, which was only clarified much later when it became known that in true hypopituitarism weight loss and emaciation are uncommon. Hormonal deficits indicative of impaired pituitary function are indeed common in anorexia nervosa, but are merely a secondary manifestation of prolonged malnutrition.

Interest in the neuroendocrinology of anorexia nervosa led to the formulation of the hypothalamic model.[6,48] From the beginning the model was aimed at explaining pathogenesis rather than aetiology; it was not considered an alternative to the psychological origin for anorexia nervosa, but a means of explaining a constellation of disturbed neural mechanisms, as follows:

1 a disordered regulation of food intake;

2 a neuroendocrine disorder affecting mainly the hypothalamic-anterior pituitary-gonadal axis;

3 a disturbance in the regulation of body temperature.

It was known that these functions all reside within the complex of hypothalamic physiology. A 'feeding centre' had been described in the lateral hypothalamus because bilateral lesions there induced self-starvation and death in rats.[49] Over the years it has become clearer that many of the disturbances could be attributed to the patients' malnutrition, as demonstrated by experimental studies in healthy young women who were asked to follow a weight-reducing diet. It was found that ovarian function is extremely sensitive to even small restrictions of caloric intake which often lead to impaired menstrual function.[50]

(b) Cerebral lesions and disturbances

Interest in the hypothalamic model was fuelled early on by clinical reports of patients diagnosed as suffering from anorexia nervosa who were later found to have cerebral lesions, especially tumours of the hypothalamus. More recently, occult intracranial tumours have been detected, masquerading as anorexia nervosa in young children.[51]

Neuroimaging studies in anorexia nervosa have led to findings suggestive of an atrophy of the brain. CT has disclosed a widening of the cerebral sulci and less frequently an enlargement of the ventricles.[52] The outer cerebrospinal fluid spaces were enlarged markedly in 36 per cent of the patients. When the CT examination was repeated after weight gain 3 months later, the widening of the sulci had disappeared in 42 per cent of the patients who had previously shown this finding. In other patients, however, the widening remained unaltered for 1 year after body weight had returned to normal.

Functional neuroimaging techniques have also been applied to research in anorexia nervosa. Regional cerebral blood flow was measured in three series of children.[53] In the majority of the children there was an above-critical difference in the regional cerebral blood flow most often between the temporal lobes but sometimes affecting other brain regions. Hypoperfusion was found on the left side in about two-thirds of the children. Follow-up scans were undertaken in children after they had returned to normal weight; the reduced regional cerebral blood flow in the temporal lobe persisted on the same side as the initial scan. The authors found that there was a significant association between reduced blood flow and impaired visuospatial ability, impaired complex visual memory, and enhanced information processing.

(c) Genetic causes

As with general psychiatric disorders, genetic and environmental factors are no longer viewed as opposing causes of anorexia nervosa, but rather as interacting with one another. Thus there may be groups of genes that determine risk of the illness and specific environmental situations that elicit or prevent the expression of these genes. There is evidence for a family aggregation of anorexia nervosa. This does not necessarily mean that the origin of the disorder is genetic because environmental factors common to the family must also be considered. In a series of 387 first-degree relatives of 97 probands with anorexia nervosa, it was found that the illness occurred in 4.1 per cent of the first-degree relatives of the anorexic probands, whereas no case was found among relatives of the controls who were probands with a primary major affective disorder, or with various non-affective disorders.[54] The authors concluded that anorexia nervosa was familial with intergenerational transmission. It was roughly eight times more common in female first-degree relatives of anorexic probands than in the general population. The absence in the relatives of probands with major affective disorder indicated the specificity in the risk of transmission of anorexia nervosa and the absence of shared familial liability with affective disorders.

The strongest argument favouring a part-genetic causation of anorexia nervosa is derived from the classical method of comparing the concordance rate of the illness in monozygotic and dyzygotic twins. Underlying the method of twin studies is the supposition that the environment for MZ twins and DZ twins is the same, or at least it does not differ in such a way as to cause greater concordance for MZ twins. Although it is known that the environments shared by MZ twins are more often similar than those shared by DZ twins, these similarities are thought not to be of aetiological relevance to anorexia nervosa.[55] The first sizeable series of twins came from the Maudsley Hospital and St George's Hospital in London and showed higher concordance rate for MZ than for DZ twins.[56] Since then the findings have been replicated and the analysis of the twin data has suggested that the liability can be broken down into three sources of variability:

a^2 Additive genetic effects	88 per cent
c^2 Common environmental effects, found to be	0
e^2 Individual-specific environmental effects.[55]	12 per cent

The individual-specific environmental effects are those which contribute to differences between the members of the twin pair. This would happen if only one member of the twin pair had suffered from an adverse effect. It is perhaps surprising that the unique environmental effects contribute to the disorder but not the twins' common environment.

Bulik has found the concept of heritability and its estimation prone to misinterpretation. There is not just one heritability estimate because this statistic varies across the population sampled and the time of the study. One estimate of heritability of anorexia nervosa gave it as 0.74 in 17-year-old female twins.[55]

It remains uncertain how the genetic vulnerability to anorexia nervosa expresses itself in terms of the pathogenesis of this disorder. One view is that this vulnerability confers instability of the homeostatic mechanisms, which normally ensure the restoration of weight after a period of weight loss. This hypothesis would explain why in western societies, where dieting behaviour is common, those who are genetically vulnerable would be more likely to develop anorexia nervosa.[57] It has also been found that MZ twins have a higher correlation for the trait of 'body dissatisfaction' than DZ twins.[55]

(d) Neurotransmitters

Since the early 1970s, the hypothalamic model of anorexia nervosa has been transformed from a consideration of anatomical 'centres' to 'systems' involving neurotransmitters. Much evidence has been presented to show that a wide range of neurotransmitters modulate feeding behaviour, and it was only a small conceptual step to suggest that some were involved in the pathophysiology of eating disorders. At first the neurotransmitters considered were mainly the monoamine systems-noradrenaline (norepinephrine), dopamine, and serotonin. In addition, opioids, the peptide cholecystokinin, and the hormones corticotrophin-releasing factor and vasopressin have also been thought to play a part in the pathogenesis of eating disorders. During recent years the main interest has been focused on the role of serotonin (5-hydroxytryptamine, 5-HT) in the control of natural appetite, especially those aspects concerning the phenomenon of satiety, mediated through a range of processes called the 'satiety cascade'.[58] There is now strong evidence that pharmacological activation of serotonin leads to an inhibition of food consumption. It was also postulated that a defect in serotonin metabolism confers a vulnerability to the development of an eating disorder.[59]

A boost to the concept of altered serotonin activity in anorexia nervosa has come from research showing that these patients while still underweight had significant reductions in cerebrospinal fluid 5-hydroxyindoleacetic acid (**5-HIAA**). The levels became normal when the patients were retested 2 months after they reached their target weight.[60] In order to test whether these findings were secondary to malnutrition, the researchers resorted to the ingenious step of studying patients after 'recovery' when they had reached normal weight. They found elevated levels of cerebrospinal fluid 5-HIAA, possibly indicating increased serotonin activity contributing to the abnormal eating behaviour which often persists in patients who have otherwise recovered.[61] The arguments against this simple model of enhanced serotonin activity as a vulnerability trait in anorexia nervosa should be briefly presented. Serotonin function was again assessed in long-term weight-restored anorectics.[62] The investigators used a dynamic neuroendocrine challenge with d-fenfluramine as a specific probe of serotonin function, which mediates the release of prolactin. If there were any persistent abnormality in serotonin function, the response to this challenge test should differ from that in normal controls. In fact, the rise in prolactin levels was very similar in former patients and normal controls. Accordingly, this study failed to support the notion of increased central serotonin function as a vulnerability trait in anorexia nervosa.[62]

Clinical features: classical anorexia nervosa (postpubertal)

The illness usually occurs in girls within a few years of the menarche so that the most common age of onset is between 14 and 18. Sometimes the onset is later in a woman who has married and had children.

By the time the patient has been referred for psychiatric treatment she is likely to have reduced her food intake and lost weight over the course of several months, and her menstrual periods will have ceased. A regular feature of the illness is its concealment and the avoidance of treatment. Even after having lost 5 to 10 kg in weight and missed several periods, the patient's opening remark is often 'there is nothing wrong with me, my parents are unduly worried'. It is only when the clinician asks direct questions that she will admit to insomnia, irritability, sensitivity to cold, and a withdrawal from contacts with her friends, including her boyfriend if she has one.

Because of this denial, it is important to enquire from a close relative, as well as the patient, about the most relevant behavioural changes.

History

History of food intake. A **food intake history** is obtained by asking the patient to recall what she has eaten on the previous day, commencing with breakfast, which is often missed. An avoidance of carbohydrate and fat-containing foods is the rule. What remains is an often stereotyped selection of vegetables and fruit. 'Diet' drinks and unsweetened fruit juice are preferred, although some patients are partial to black coffee. It is the mother who will indicate that her daughter finds ways of avoiding meals, preferring to prepare her own food and withdrawing into her bedroom to eat it.

Weight history. The patient is usually willing to provide an accurate **weight history**. She may try to conceal her optimum weight before her decision to 'diet', but she is likely to be objective about her current weight, if only to express pride in the degree of 'self-control' she has exerted. The clinician then has an opportunity to enquire into her 'desired' weight by simply asking what weight she would be willing to return to. Her answer will betray a determination to maintain a suboptimal weight.

History of exercising. A **history of exercising** should be taken. Again, the patient is likely to conceal the fact that she walks long distances to school or to work rather than use public transport. She may also cycle vigorously or attend aerobic classes. A parent may report that his or her daughter is running on the spot or performing press-ups in the privacy of her bedroom, judging from the noise that can be heard. The amount of exercising may be grossly excessive, with the patient indulging in brisk walks or jogging even in the presence of painful knees or ankles due to soft tissue injuries.

Additional harmful behaviours, which should be enquired into include self-induced vomiting, purgative abuse, and self-injury. Vomiting and purgative abuse are similar to the behaviours that occur in bulimia nervosa (see Chapter 4.10.2). In anorexia

nervosa they may occur without the prelude of overeating and the patient's motive is simply to accelerate weight loss. The laxative abuse is often at the end of the day. The favourite laxatives in the United Kingdom are Senokot and Dulcolax, and the patient is likely to take them in increasing quantities to achieve the wanted effect as tolerance develops. Self-injury should also be enquired into, and the skin of the wrists and forearms inspected for scratches or cuts with sharp instruments.

Menstrual history. The patient may not volunteer that her periods ceased soon after commencing the weight-reducing diet. On the other hand she may admit that she is relieved that her periods have stopped as she found them to be a nuisance or unpleasant.

(a) The patient's mental state

(i) Specific psychopathology

Several near-synonyms have been used to describe the specific attitude detectable in the patient who systematically avoids fatness: a 'disturbance of body image',[5] a 'weight phobia',[29] or a 'fear of fatness'.[6] Magersucht, or seeking after thinness, was a term applied in the older German literature. The patient will express a sensitivity about certain parts of her body, especially her stomach, thighs, and hips. Not only is she likely to assert that fatness makes her unattractive, but she may add that it is a shameful condition betraying greed and social failure. These distorted attitudes generally amount to overvalued ideas rather than delusions. Occasionally, however, a patient may be frankly deluded, such as one young woman who believed that her low weight was due to thin bones and that fatness was still evident on the surface of her body.

Studies have demonstrated that wasted patients, when asked to estimate their body size, see themselves as wider and fatter than they actually are.[63] Since these early observations, numerous perceptual studies have been undertaken and the conclusion drawn that anorexic patients overestimate their body width more often than normal controls. These distorted attitudes are often associated with a negative affect, so that the disturbance might be viewed as one of 'body disparagement'.[64]

The patient's dread of fatness is so common that it is pathognomonic of anorexia nervosa. There are, however, exceptions. Sometimes a patient may simply deny these faulty attitudes. Another exception is the occurrence of anorexia nervosa in eastern countries where thinness is not generally admired (e.g. Hong Kong and India). The imposition of fear of fatness as a diagnostic criterion on patients from a different culture, where slimness is not valued, amounts to a failure to understand the illness in the context of its culture.[65] An appropriate solution, proposed by Blake Woodside[66] is to accept diagnostic criteria where the psychopathology includes culture-specific symptoms which will differ for western, Chinese, and Indian ethnic groups.

(ii) Denial

Denial in anorexia nervosa is sometimes included within diagnostic criteria of the disorder, for example DSM-IV. Vandereycken[67] has written a scholarly treatize on the subject. He has pointed out that the term 'denial' is used with apparently simplicity in clinical practice, but this stands in contrast to its intriguing complexity. It is a multi-layered concept, which is difficult to measure. Denial is not a static condition but fluctuates over time. A crucial element in its assessment is the inherent conflict of perception between the patient and the clinician. The patient carries out distortions by omissions, concealments, or misrepresentations. She often denies hunger and fatigue and appears not to accept that she is thin. She shows a lack of concern for the physical and psychosocial sequelae of being underweight and may even deny the danger of her condition.

Vandereycken has proposed two categories of denial:

1 Unintentional denial (e.g. the patient's way to improve her self-esteem and preserve her sense of identity).

2 A deliberate denial including the pretence of being healthy and avoiding the treatment others want her to accept. Strong denial is often accompanied by the avoidance of treatment, but the refusal of help is not entirely due to the patient's lack of recognition of her problems.

(iii) Depression

Depression of varying severity, including a major depressive disorder, is common. The patients express guilt after eating, adding that they do not deserve food. A high rate of depression (42 per cent) was found at presentation in one study[21] and the lifetime history of depression in follow-up studies may be as high as 68 per cent.

(iv) Obsessive–compulsive features

The patients frequently eat in a ritualistic way, for example restricting their food intake to a narrow range of foods which experience tells them are 'safe' because they will not lead to weight gain. There is often a compulsive need to count the daily caloric intake. One patient rejected prescribed vitamin tablets in case they contained 'calories'. The frequency of obsessive–compulsive disorder in anorexia nervosa was found to be 22 per cent in a clinical series.[21] In studies using structured clinical interviews the frequency ranged from 25 to 70 per cent.

(v) Neuropsychological deficits

These deficits are seldom detected clinically, and an emaciated patient may pursue studies and obtain surprisingly good examination marks. On the other hand, neuropsychological testing has revealed deficits in attention, and impairment of memory and visuospatial function.[59]

(b) The endocrine disorder

(i) The impairment of hypothalamic-pituitary-gonadal function

Amenorrhoea is an early symptom of anorexia nervosa and in a minority of patients may even precede the onset of weight loss. Amenorrhoea is an almost necessary criterion for the diagnosis of anorexia nervosa. An exception is when a patient takes a contraceptive pill, which replaces the hormonal deficit and may lead her to say she still has her periods.

Generally, when the patient is undernourished, levels of gonadotrophins and oestrogens in the blood are found to be low or undetectable. Not only do malnourished patients show low blood levels of luteinizing hormone and follicle-stimulating hormone, but the secretion patterns of these pituitary hormones regress to a phase of earlier development. For example, severely wasted patients display an infantile luteinizing hormone secretory pattern with a lack of major fluctuations over the course of 24 h. With some degree of weight gain a pubertal secretory pattern appears, consisting first of a sleep-dependent increase of luteinizing hormone at night, and later displaying more frequent fluctuations.[68]

When a patient is still malnourished the ultrasound pelvic examination will reveal that ovarian volume is much smaller than in normal women.[69] Three stages can be discerned in the appearance of the ovaries as the patient gains weight:

1 small amorphous ovaries;

2 multifollicular ovaries (with cysts 3–9 mm in diameter);

3 dominant follicle (10 mm or more in diameter).

At the same time there is a corresponding return of hormonal secretion; follicle-stimulating hormone appears first, followed by luteinizing hormone and finally oestradiol which leads to enlargement of the uterus.[70]

These abnormalities signify that the patient is infertile and remains so until endocrine function recovers. Pregnancy occasionally occurs as the patient is still underweight and improving but before the appearance of the first menstrual period.[71] The pregnancy carries a risk of poor foetal growth during the first trimester, albeit with some 'catch-up' growth during the neonatal period.[72] Occasionally an underweight anorexic patient may seek treatment at an infertility clinic. Treatment with gonadotrophin-releasing hormone may restore fertility, but this practice has been severely criticized.[73]

(ii) Hypothalamic-pituitary-adrenal and hypothalamic-pituitary-thyroid axes

The emaciated anorexic patient shows an increased 24-h plasma cortisol level which returns to normal with a minimal increase in weight, as it is the nutritional intake that is critical.[74]

Reduced tri-iodothyronine (T_3) levels are linked to a reduction in energy expenditure during starvation and are adaptive in nature, so that treatment with L-thyroxin is not appropriate in anorexia nervosa.[68]

(c) Weight loss and malnutrition

(i) Body weight

The severity of weight loss may be recorded as follows:

1 As a percentage of an 'average' body weight to be found in tables for normal populations according to age and height (e.g. the *Metropolitan Life Insurance Tables*).

2 As a percentage drop from the patient's 'healthy' weight before the onset of her illness.

3 Using the Quetelet body mass index

$$\text{BMI} = \frac{\text{weight (kg)}}{(\text{height (m)})^2}$$

A BMI between 20 and 25 is regarded as healthy. A BMI of 14 or less indicates a need for hospitalization. A BMI between 10 and 12 represents the lower limit of human survival.

(ii) Physical examination

Wasting is variable but may be extremely severe, resulting in a skull-like appearance of the head, stick-like limbs, and flat breasts, buttocks, and abdomen. The hands and feet feel cold and readily turn blue in winter. The skin is dry with an excess of downy hair (lanugo) covering the cheeks, the nape of the neck, the forearms, and the legs. Heartbeat is slow (50–60 beats/min) and the blood pressure is low (e.g. 90/60 mmHg) with orthostatic lability. During the routine physical examination muscle power should be tested to detect proximal myopathy, as explained below.

Differential diagnosis of classical anorexia nervosa

The diagnosis of anorexia nervosa is usually straightforward, especially as the modern diagnostic criteria are objective. Wasting diseases such as inflammatory bowel disease (Crohn's disease or ulcerative colitis), thyrotoxicosis, and diabetes mellitus may sometimes be mistaken for anorexia nervosa, but they can be identified through specific investigations. Occasionally there is an interaction between such a medical illness and anorexia nervosa, when a patient wishes to perpetuate the weight loss caused by the former. Rarely, anorexia nervosa may be mimicked by a cerebral tumour altering the function of the hypothalamus.

In older patients there may be diagnostic difficulty with a major depressive or schizophrenic illness. The weight loss in a severe depressive illness results from loss of appetite and the patient's belief that she does not deserve food. A schizophrenic patient may avoid food because of paranoid delusions of being poisoned.

Complications of malnutrition

The complications, which are part of the illness, such as amenorrhoea due to failure of the hypothalamic-pituitary-gonadal axis, have already been discussed. The range of medical complications has been extensively reviewed,[74] with recommendations for the investigations needed on presentation of the patient.

(a) Fluid and electrolyte disturbances and cardiovascular complications

Patients who induce vomiting, abuse laxatives, or take diuretics are likely to experience dehydration and various electrolyte disturbances.

(i) Self-induced vomiting

Loss of gastric acid leads to a metabolic alkalosis and hypokalaemia. A low serum level of potassium may lead to cardiac conduction defects and arrhythmias, skeletal muscle weakness, and renal tubular dysfunction.

(ii) Laxative abuse

The abuse of laxatives is likely to cause dehydration, metabolic acidosis, hyponatraemia, and hypokalaemia.

(iii) The use of diuretics

The use of diuretics such as the thiazide group gives rise to low serum sodium levels which in turn cause fatigue and general weakness. A level below 120 mmol/l may lead to coma. The patient often justifies the use of a diuretic as a treatment for swollen ankles or even 'bloating' of the stomach, which she misinterprets as being due to fluid retention.

(iv) Impaired renal function

This may be caused by different mechanisms: pre-renal failure due to dehydration or hypokalaemic nephropathy.

(b) Peripheral oedema

Fluid retention leading to oedema occurs frequently in patients with anorexia nervosa. It is important to distinguish between 'benign' oedema, which often occurs during the course of effective refeeding, and oedema from other causes, which may have serious consequences.

(i) 'Benign' oedema

'Benign' oedema may occur during inpatient treatment. If the patient accepts a high-caloric intake, fluid retention is the rule as shown by a steep upward rise in the weight curve.

(ii) Famine oedema

If oedema is detected when the patient first presents clinically, or if it develops without a preceding improvement in food intake, the underlying mechanisms should be carefully investigated in order to avoid the risks of congestive cardiac failure and pulmonary oedema. By the time peripheral oedema is detectable, the amount of fluid retained in the body contributes several kilograms to the patient's weight, thereby concealing the true loss of weight. It is a misconception that peripheral oedema is usually due to a lowering of plasma albumen, as this is not necessarily so in anorexia nervosa. Therefore the clinician must look for other reasons.

1 A fall in interstitial fluid pressure has been proposed whereby water seeps from the blood into the interstitial spaces[75] whereas the total exchangeable sodium is likely to remain within the normal range.

2 'Rebound' oedema following hyponatraemia due to abuse of laxatives or diuretics.

3 Wet beri–beri: vitamin deficiency in anorexia nervosa is uncommon because many patients continue to eat vegetables and fruit. Nevertheless a lack of vitamin B_1 (thiamine) may occur in patients who eat a stereotyped diet, and Wernicke's encephalopathy may follow.[76] It may also lead to congestive cardiac failure with severe oedema from a nutritional cardiomyopathy, precipitated by refeeding without thiamine.

4 Congestive heart failure and pulmonary oedema may occur as a result of general undernutrition leading to a decreased cardiac mass, cardiac output, and volume. **Death may be caused by an injudicious intravenous infusion of fluids**.

(c) Metabolic disturbances

(i) Hypoglycaemia

Severe hypoglycaemia with plasma glucose levels as low as 1.0 mmol/l is recognized as a cause of death. Hypoglycaemia may go unrecognized, as a lack of sympathetic nervous response may mask the classical symptoms and signs.

(ii) Hypercarotenaemia

This is a benign condition causing an orange–yellow colouration of the skin of the palms, soles, forearms, and the region of the nose. It is partly due to the consumption of large amounts of foods rich in carotenoids, especially carrots, tomatoes, and the green outer leaves of vegetables.

(d) Myopathy

Weakness of specific muscle groups is common and is due to severe protein-energy malnutrition. There is a 'proximal' myopathy, affecting the musculature of the pelvic girdle and the shoulder girdle. The patient first notices an increasing difficulty in climbing stairs. Weakness may also affect the muscles of the head and neck and the face. When the patient is asked to rise from a sitting position, she will tend to push herself up using her hands and arms. She also has difficulty in rising unaided from a squatting position.[78]

There are no characteristic abnormalities in blood chemistry, although creatine kinase and liver enzymes may be elevated and the activity of the enzyme carnosinase may be reduced. Myopathic changes are consistently present in muscle biopsy specimens. Histology reveals the 'chequerboard' distribution of muscle fibres with a selective type 2 fibre atrophy. Electron microscopy reveals the presence of strikingly abundant glycogen granules between the myofibrils and under the sarcolemma.[78]

The detection of myopathy is a clear indication for the patient's admission to hospital and a refeeding programme. After a weight gain of a few kilograms her muscle strength will begin to return and a complete recovery is the rule.

(e) Osteoporosis

A high proportion of patients with anorexia nervosa risk developing osteoporosis and consequent pathological fractures. Significant reduction in bone mineral density of the femoral neck was found in all 20 patients with anorexia nervosa of 6 years or more duration.[79] The favourite method of measuring bone density in the lumbar spine and hip is by dual-energy X-ray absorptiometry. A measurement for all patients with anorexia nervosa of 2 years duration or more is recommended.[74, 79] It is difficult to disentangle the harmful effect of the nutritional deficiency itself from the associated oestrogen deficiency. The pathogenesis of osteoporosis in anorexia nervosa differs from that in postmenopausal women. In anorexia nervosa the nutritional deficiency (including a lack of calcium and vitamin D) leads to a low rate of the recycling of bone through bone formation and resorption, but the balance is disturbed with a relative increase in bone resorption.

The evidence favours the likelihood of improvement of bone density with nutritional recovery and resumption of menstruation.[79] There is much uncertainty about the best treatment. Although patients are often automatically prescribed oestrogen replacement, the only controlled trial undertaken so far indicated that oestrogen was only effective in severely underweight anorexic patients.[81] Indeed, it has been argued that oestrogen replacement could be harmful in some patients (children with premenarchal onset) and unnecessary in others with an illness of less than 3 years duration.[82] Instead the emphasis should be on encouraging weight gain, with the possible addition of calcium and vitamin D. Older patients with a poor prognosis (e.g. who have been ill over 10 years) might benefit from oestrogen replacement. There is little evidence for the use of other medication such as biphosphonates.

(f) Disturbed temperature regulation

Disturbances in the control of temperature are evident from clinical observations; the patients frequently complain of feeling cold, and in the winter they have cold and blue extremities and suffer from chilblains. In the severely malnourished patient hypothermia may be a cause of death. Severe malnutrition is accompanied by a low central body temperature, presumably because of a low metabolic rate. Ingestion of a high-calorie meal can cause a significant increase in the central body temperature[83] with complaints of heat in the periphery and sweating after food.

(g) Haematological changes

Anorexic patients may develop a significantly lowered haemoglobin level and a reduced haematocrit and white cell count, with a relative decrease in neutrophil leucocytes and an increase in lymphocytes.[84] The anaemia is usually moderate and is normocytic normochromic in type. The mechanism is that of starvation-induced bone marrow hypoplasia. Only occasionally is there an iron deficiency. The anaemia gradually becomes corrected with weight gain. A low platelet count in the blood has also been observed and there may be an associated thrombocytopenic purpura.

(h) Complications arising during rapid refeeding

(i) Acute dilatation of the stomach

This complication has been described in anorexia nervosa during the course of refeeding.[85] The patient develops copious vomiting, upper abdominal pain, distension of the upper abdomen, and rapid dehydration. Treatment is by continuous gastric aspiration, and this complication is one of the rare indications for intravenous infusions of glucose and saline. Gastric dilation is best prevented by avoiding a food intake above 1500 cal daily during the first week of refeeding.

(ii) Hypophosphataemia

When the illness has been protracted and has led to severe emaciation, the body stores of phosphate become depleted. The fall in serum phosphate is aggravated during refeeding, especially parenteral feeding.[86] Levels of phosphate may fall as low as 0.2 mmol/l. Clinical features are cardiac irregularities with a prolonged QT interval, impairment of consciousness, and delirium. Treatment is with oral phosphates rather than by intravenous administration.

Anorexic mothers as parents

A patient who is improving may conceive despite having a suboptimal weight and still not menstruating.[71] A mother may also develop the illness after having borne children. In a series of eight mothers, 9 out of 13 of their children suffered from food deprivation, identified by reductions in weight for age and in height for age as shown on Tanner–Whitehouse charts.[87] The anorexic mothers had no intention of abusing the children and indeed were affectionate towards them. **They simply extended their abnormal concern with body size to their own children**. They adopted different ways to ration their children's food intake according to their age. They might prolong breast feeding, dilute the bottle feeds, reduce the amount of food available in the home, confine eating to meal times, forbid the consumption of sweets, and prevent others giving them food. An important part of the management is to anticipate the risk to any children and conduct tactful enquiries. A whole-family approach should be adopted, focusing on the children's needs for food. The children should be followed up to ensure that they gain weight and catch-up growth is established.[87]

Clinical features: anorexia nervosa of early onset (premenarchal)

It would be too arbitrary to define an early onset by age limits such as an onset from 8 to 16 years. A more meaningful frame of reference is the onset in relation to the stage of puberty reached by the child.[88] Because puberty is a complex developmental process spanning 2 to 3 years, it is best to name as 'premenarchal' the illness which commences some time after the first signs of puberty and before its completion as shown by the first menstrual period. In true prepubertal anorexia nervosa the illness begins even earlier, before the very first signs of puberty. Postpubertal anorexia nervosa is when the illness commences after menstruation has been established.

There is much similarity between the clinical features of an illness of early onset and one which is postpubertal. However, there are two important differences. The first is the potential for the illness to interfere with the child's pubertal development. The second is the heart-rendering impact on the child's family. It follows that the management of the family is of supreme importance, and the clinician should be prepared for parental reactions, which may detract from a rational plan of treatment.

The clinical presentation is variable. Often there are precipitating events such as a family bereavement or a physical illness leading to weight loss. The onset is likely to be insidious,[18] with the parents noticing nothing amiss except for non-specific features. Symptoms of depression and irritability are common.[14,77] Some children cannot describe feelings of depression, but tearfulness may be obvious. They withdraw from friends and may refuse to go to school. Others express ideas of being underserving of love or food. Increasingly these youngsters have been found to injure themselves, especially by scratching with their nails or cutting the skin of the wrists and forearms with sharp objects, and occasionally by knocking and bruising the head. In a severe depression the child may say she hears voices calling her 'bad', but further questioning indicates that these are not true hallucinations but vivid expressions of her own thoughts. Another common presentation is with bodily symptoms, especially headaches, abdominal discomfort, and a wide range of gastrointestinal symptoms, which inevitably elicit physical investigations.

At some stage, however, the parents observe that their child is avoiding food and is reluctant to eat at normal meal times. She resorts to deviousness and secrecy. The omission of school meals often goes undetected. Eventually it is noticed that she has become thinner and may have lost a great deal of weight. Resistance is met when attempts are made to reverse the loss of weight. Even a young child may become preoccupied with the caloric values of foods and adopt additional methods of inducing weight loss. The child is likely to exercise excessively—jogging, walking, or cycling long distances. An attempt to reduce the excess activity may lead to solitary exercising in the bedroom, including press-ups or running on the spot. Other patients may induce vomiting or take laxatives even after small meals, but overeating typical of bulimia nervosa is rare in young children.[77]

(a) Weight loss

Because of the early onset while the child should still be growing in stature, there is a failure to gain the weight, which normally accompanies the growth spurt. Later there occurs an actual loss of weight and, because the child has not yet reached her optimum weight, a very low weight indeed may result—25 kg or even less. Symptoms of malnutrition ensue including tiredness, constipation, and sensitivity to cold with cold extremities.

(b) The psychological disorder

Even younger children are likely to disclose that they are fearful of becoming fat, a disturbance similar to the overvalued idea held by older patients. A minority of patients will disclose their reluctance to develop personal, sexual, or social maturity, in a manner, which fits into Crisp's model (see p. 796 of this chapter). A few may express reluctance to have menstrual periods. A girl may say she is indifferent whether she menstruates or not, but would like to develop breasts like other girls in her class. The reluctance to 'grow up' may be expressed in social terms, with the patient saying that she could not imagine herself ever leaving home or her mother. On the other hand most girls are keen to reach a normal stature.

Severe depression was found in 69 per cent of youngsters in the Göteborg study.[14] In the same series one-third of the patients had an obsessive–compulsive personality disorder, and 8 per cent developed hand-washing and other compulsions.

(c) Physical examination

Physical examination will reveal a varying degree of wasting, affecting the limbs, the abdomen, the buttocks, and the facial appearance. The extremities are blue and cold, and ischaemic changes may lead to gangrene affecting the toes. Other physical changes are similar to those in the adult, with the addition of a delay in puberty.

(d) Delay or arrest of puberty

The illness may adversely affect pubertal development depending on its time of onset. If the onset is truly prepubertal the child will not yet have shown the first signs of puberty, such as the appearance of pubic hair and breast buds. When the illness begins during the course of puberty these early signs may have appeared, but the breasts will show early growth only (Tanner stages 1 or 2), and the child will not yet have menstruated. The effects on pubertal development can be divided into those affecting growth in stature, breast development, and menstrual function.[88]

(i) Growth in stature

Growth in stature may have become arrested. In a series of 20 patients with a premenarchal onset, only two of them had reached the 50th centile in height. With successful treatment and weight gain, catch-up growth of 2 to 5 cm may be achieved, but only in patients aged 17 or less.[88]

(ii) Breast development

Breast development: in the same series only six patients had normal breasts and as many as 10 had infantile breasts. After prolonged weight gain, eight of the 14 patients showed a considerable response in breast growth.[88]

(iii) Menstrual function

Menstrual function: Primary amenorrhoea is the rule. In the series already referred to only four of the 20 patients had menstruated by the age of 16 years. A further three began their periods between 16 and 18 years of age, and four at ages ranging from 18 to 25. The remaining patients had prolonged amenorrhoea.

The above series consisted of selected patients in whom pubertal delay was severe, whereas marked pubertal delay has seldom been reported in other series in which only delayed menstruation has been remarked upon.

A young boy who develops anorexia nervosa is also likely to become preoccupied with fatness and accordingly avoids food in order to lose weight. The endocrine disorder in the male similarly consists of a disturbance of the hypothalamic-pituitary-gonadal axis. With a prepubertal onset, the penis and scrotum remain infantile, there is only a scanty growth of pubic and facial hair, and the voice may not break.

Special investigations

Pelvic ultrasound monitoring of the ovaries and uterus is a useful method of ascertaining regression and recovery in children with anorexia nervosa.[89] On first testing, and in the presence of low weight and amenorrhoea, ovarian volume and uterine volume were found to be reduced in comparison with normal pubertal girls. In the latter the normal range of ovarian volume is 3.95 ± 1.7 cm^3 and uterine volume is 14.8 ± 7.6 cm^3. On retesting the patients 18 months after the first scan those who were menstruating showed significantly larger ovarian and uterine volumes than those with amenorrhoea. The authors concluded that for ovarian and uterine maturity to occur it is necessary to achieve a mean weight of 48 kg and a mean weight-to-height ratio of 96.5 per cent. They found that pelvic ultrasound scanning helped to motivate the children to accept a higher body weight.

Differential diagnosis of early-onset anorexia nervosa

Frequent bodily complaints, loss of weight, and abnormalities of growth lead these children to be referred to paediatricians for special investigations. It has been proposed that young anorexic boys should be investigated by means of neuroimaging, so as not to miss occult intracranial tumours.[32] The diagnosis of anorexia nervosa should be distinguished from atypical eating disorders in childhood.[77]

Pervasive refusal syndrome is characterized by a child refusing to eat, drink, walk, talk, or take care of herself. Anxiety, phobic responses, and depression are also present.

Selective eating is the term applied to a child who restricts food intake to two or three different foods, such as biscuits, crisps, or potatoes, but usually remains in good health.

Food avoidance emotional disorder. this condition is one in which food avoidance is attributable to an emotional disturbance in the absence of a dread of fatness, a necessary criterion for the diagnosis of anorexia nervosa.

Food fads and food refusal. they are usually intermittent and physical health is not compromised.

Classification

ICD-10[80] (including patients with premenarchal onset[88])

1 Body weight is maintained at least 15 per cent below that expected for health. Prepubertal children may fail to gain the weight expected during the growth spurt. Weight loss is caused by avoidance of 'fattening' foods, possibly with the addition of self-induced vomiting, purgative abuse, excessive exercise, or the use of appetite suppressants.

2 The patient holds the overvalued idea of a dread of fatness and keeps her weight below a self-imposed threshold.

3 There is an endocrine failure manifest in women as amenorrhoea and in men as a loss of sexual interest and potency.

4 If the onset is premenarchal, the sequence of pubertal events is delayed: growth is arrested; in girls the breasts do not develop and there is a primary amenorrhoea; in boys the genitals remain juvenile.

DSM-IV[90]

1 Refusal to maintain body weight above the minimally normal weight for age and height (e.g. weight less than 85 per cent of that expected).

2 Intense fear of gaining weight or becoming fat, even though underweight.

3 Disturbance in the way in which one's body weight or shape is experienced.

4 In postmenarchal females, there is amenorrhoea of at least three menstrual cycles.

DSM-IV subdivides anorexia nervosa as follows.

1 Restricting type, without regular binge-eating or 'purging' behaviour (in the United States this term includes self-induced vomiting and the misuse of diuretics as well as laxatives).

2 Binge-eating/purging type.

Anorexia nervosa in males

The reduced frequency of anorexia nervosa in the male might lead one to surmise that the disorder is likely to differ between the sexes in its aetiology, clinical manifestations, and prognosis. Yet, there are remarkable similarities between the sexes as regards the age of onset and the specific features of the psychopathology.[91–93] For example, the male patients tended to select a diet which was low in fattening foods and resorted to subterfuges to dispose of food, such as self-induced vomiting and purging, and strenuous exercising. They expressed a fear of fatness and considered themselves overweight, even when they were thin.

Of course there are fundamental biological differences which inevitably alter the manifestations of the endocrine disorder in the male and, to a lesser extent, the nutritional disorder. Testicular function, as gauged by the urinary output of testosterone, is disturbed in male patients when they are emaciated. Refeeding leads to a partial correction of this abnormality.[91] The body composition of the mature male differs from that in the female; he has a lower reserve of adipose tissue so that protein depletion occurs more rapidly when he loses weight.

The relative resistance of the male against developing anorexia nervosa remains a mystery. It is even unclear whether the sex difference is likely to be due to biomedical factors or psychosocial differences. It has been suggested that young females often become preoccupied with 'fatness' because of its reproductive, biological, and social significance, whereas young males are more concerned with their musculature and its significance for strength, dominance, and masculinity.[92] These differences are linked with the frequency of dieting among adolescent girls and its rarity in boys.[94]

In a series of male patients with 'primary' anorexia nervosa most of the patients reported problems with sexuality.[95] Sexual anxieties had been present with respect to heterosexual as well as homosexual behaviour. One quarter admitted homosexual tendencies. Almost all were relieved by the loss of libido following weight loss. The authors concluded that males with atypical gender role behaviour had an increased risk of developing anorexia nervosa in adolescence.

There are only a few follow-up studies of anorexia nervosa in males. In one impressive study, 27 patients were followed up for a minimum of 2 years and a mean of 8 years.[96] Expressed in terms of the Morgan–Russell categories of general outcome, a good outcome was found in 44 per cent, an intermediate outcome in 26 per cent, a poor outcome in 30 per cent, and no deaths. Only a few predictors of outcome were identified: disturbed relationships with a parent in childhood led to a poor outcome, and the occurrence of previous sexual activity was associated with a good outcome. The outcome in males was remarkably similar to that in females.[94]

Course and prognosis

The natural outcome is defined as the long-term results of the illness or disease process. The clinical prognosis is the difficult task of forecasting the future course and final outcome of the illness in an individual patient.

Outcomes from follow-up studies in anorexia nervosa

There have been comprehensive appraisals of follow-up studies in anorexia nervosa[97] which have put forward criteria for the near-perfect follow-up study, which in practice are seldom fully met. Among these criteria are precision in the diagnostic features, the use of standardized interviews, 100 per cent success in tracing the patients, and a sufficiently long follow-up to determine eventual outcome. An arbitrary interval of at least 4 years was previously set[21] and most recent studies have adhered to this recommendation. Several groups of investigators have adopted measures of outcome based on the Morgan–Russell scales.[98] Their use gives rise to three possible categories of general outcome based on body weight and menstrual function: 'good', 'medium', and 'poor'.

In a review comparing three British studies and one Swedish study,[99] each with a mean follow-up of 5 to 6 years, it was found that the patients treated in Bristol had a better outcome than those treated at the Maudsley Hospital in London. The explanation is one of selection bias already mentioned: the Maudsley patients had all required inpatient treatment, whereas in the Bristol series the majority were outpatients. The third British study, from St George's Hospital, London, showed a quality of outcome intermediate between the other two.

The Swedish study was extended by two later follow-ups at 15 and 33 years. On the one hand, the percentage of good outcomes gradually increased and the percentage of poor outcomes diminished. On the other hand, the death rates increased with time; after 33 years the total mortality from anorexia nervosa or suicide was 18 per cent. Slow recovery was the rule: only 29 per cent of patients recovered in less than 3 years of illness, another 35 per cent within 3 to 6 years, and the remainder took much longer with recovery after 12 years being rare.[99] The Maudsley series of patients was also extended to a mean follow-up of 20 years. There was a reduction in the percentage of patients with a poor outcome, but an increase in the mortality rate close to that of the Swedish study.[100]

Prognostic predictors of outcome

In the long-term Maudsley study a poorer outcome was predicted by an older age of onset, a longer duration of illness, the presence of neurotic traits in childhood, personality problems, and the occurrence of relationship difficulties within the family.[100]

Comparison of mortality rates

In a review of 42 studies the aggregate annual mortality rate from anorexia nervosa was found to reach 0.56 per cent on average.[101] Complications of anorexia nervosa accounted for 54 per cent of deaths, suicide for 27 per cent, and other causes for 19 per cent.

A fair measure of consistency has been found in different parts of the world, especially when allowance is made for selection biases: in Denmark 0.5 per cent per annum (younger patients), in Sweden 0.75 per cent per annum, in the United States 0.66 per cent per annum, and in the United Kingdom 0.75 per cent per annum.[100]

Follow-up studies in early-onset anorexia nervosa

In a series of 60 patients in Berlin with a mean age of onset of 14.7, followed up for at least 4 years, recovery occurred in 68 per cent of patients but there was a 6.6 per cent mortality over a mean of 4.8 years,[102] a higher rate than was found in the Danish study already mentioned. A study of teenage-onset anorexia nervosa had the advantage of community screening and the use of a comparison group. At a 10-year follow-up the outcome was good in 43 per cent, intermediate in 29 per cent and poor in 20 per cent. There were no deaths. Strikingly there was a poor psychosocial outcome in half the anorexic patients with high dependence on state benefits.[103]

Conclusions

Theander[104] has provided good advice and useful definitions:

> Clinical experience has repeatedly shown that patients with anorexia nervosa can recover after a period of illness amounting to more than 10 years.

A distinction should be made between a 'chronic' illness and a long-standing illness. The word 'chronic' should be restricted to a continuous illness of more than 15 years' duration because up to then the patient is still capable of a full recovery. The term 'protracted' is preferable for an illness of prolonged duration but not as long as 15 years.

Treatment: review of evidence

In this review the evidence underpinning treatments for anorexia nervosa will be presented. Whenever possible the clinician should use a treatment whose efficacy has been established, but he will seldom be able to confine himself to such treatments. For example, in the case of the severely malnourished patient who presents as an emergency an admission to the specialized eating disorder unit may be the first necessary step, even though it has sometimes been argued that the evidence underpinning inpatient treatment is limited. In the majority of patients presenting with only moderate weight loss it is possible to select a treatment with a higher level of evidence for its efficacy. Such a treatment is family therapy, the treatment of choice in an adolescent patient.

Some cautionary words are called for on the subject of evidence-based psychiatry. For anorexia nervosa, there is no truly specific treatment in the sense of effecting a long-term 'cure'. The choice of treatment should be determined by the needs of the patient at the time when she presents for treatment and will vary according to the severity of the illness and the stage it has reached. It is wise to take into account the patient's age, her maturity, and her personality. A given treatment may be indicated for a short period only at the end of which the therapeutic aims need to be reviewed. Such a selective approach has sometimes been called the 'stepped care' approach but this concept can be misleading. The steps may not be in the same direction. They may go down as well as up. For example, a patient may be in urgent need of refeeding and if she makes good progress the next stage would consist of consolidating her improvement. If she is unfortunate, however, earlier priorities will reassert themselves.

A tendency to decry the benefits of treatment in anorexia nervosa first appeared in the early 1990s, and in more recent years cries of despair have become ever shriller. To some extent this is a side effect of the treatment evaluation industry which has unintentionally caused confusion between the usefulness of treatments on the one hand and the solid scientific evidence for their efficacy on the other. In an ideal world the two would be the same. In an illness such as anorexia nervosa they often are not, as will be illustrated further on.

A key member of the treatment evaluation industry in the United Kingdom is **the National Institute for Clinical Excellence** (NICE).[105] NICE has developed a hierarchy for the strength of the evidence underpinning various treatments, as described in Chapter 1.10. I is the strongest level of evidence, followed by Ia, IIa, IIb, III, and IV, the last being the lowest level. Anorexia nervosa is an area of care, which particularly requires Clinical Practice guidelines because of substantial dangers to life and physical health and some clinical uncertainties. So far NICE has failed to provide the required 'statements to assist practitioner decisions about appropriate health care for specific clinical circumstances'. NICE has not recognized the solid evidence favouring family therapy in adolescents (according it only Grade B in an ill-defined set of Grades A, B, and C for clinical guidelines). Family therapy should be regraded as A. NICE has also failed to identify the specific value of a specialized eating disorder service in patients with a risk to physical health and life. This should have been graded at least as B, or as (I) using the system of clinical recommendations described in Chapter 1.10. The correction of severe malnutrition requires a specialized eating disorder service rather than the involvement of a physician and medical unit (proposed by NICE) whose treatment is usually restricted to tube-feeding. NICE has also failed to recognize the value of other treatments, according them only a Grade C, implying there have been no good clinical studies.

The randomized controlled trial (RCT) has been idealized as the 'gold standard' for the level of evidence. The absence of such evidence for a given treatment does *not* signify that this treatment lacks efficacy. Some colleagues write and speak as if treatments other than those confirmed by RCT are not worthy of consideration. In truth they may well be effective, but the evidence for this is hard to come by. An effect of concentrating on RCTs and ignoring sound clinical evidence, is the conclusion reached by one authority that evidence-based treatment of anorexia nervosa is 'barely' possible. This is to ignore the fact that there are essential treatments for which evaluation by RCT has not yet been possible, because of insuperable ethical obstacles or limited patient compliance.

RCTs are particularly problematic in the assessment of treatments for patients with anorexia nervosa. For example, the NICE guidelines do not mention the role of a specialized eating disorder service, although there are oblique references to 'professionals with specialist experience of eating disorders' and 'a place for inpatient management'. Specialized eating disorder units (EDUs) long ago graduated to providing a broad service not confined to inpatient treatment but including daycare and outpatient treatment. It is important to evaluate the benefits of a specialist eating disorder service, and this is not mentioned in the NICE Guidelines. There has been only one study randomizing low weight patients to inpatient or outpatient treatments,[96] as the difficulties are obvious.

A balanced view needs to be taken towards the role of RCTs in the evaluation of treatments in anorexia nervosa. There has recently been a swing of the pendulum in the reverse direction. Some researchers have reached the surprising conclusion that RCTs in anorexia nervosa are so difficult that they should be considered as 'premature' in view of their cost and the failure of some of hem.[106]

An important reason why researchers may fail to complete a satisfactory RCT is the patient's behaviour in the course of the trial. The anorexic patient's resistance to treatment may extend to sabotaging the basic requirements of the treatment trial itself, e.g. compliance with the treatment and the follow-up assessments. These problems are difficult to eliminate entirely. Nevertheless RCTs are still possible if the researchers are assiduous in following-up their patients. Evaluators of research trials should see that the perfect trial in anorexia nervosa is a contradiction in terms. The pessimists favouring the abandonment of RCTs should remember that new ideas can emerge unexpectedly during the course of a trial, thus pointing the researcher towards fresh treatment approaches capable of being retested at a later date.

In the remainder of this section levels of evidence for the efficacy of treatments will be discussed, even when they are not considered by NICE.

Family therapy

Family therapy is the main treatment to have been carefully evaluated in anorexia nervosa, having now been tested in a number of randomized controlled trials. Any one of these trials has its imperfections, but taken as a body of evolving evidence, family therapy should be seen as the method of choice for treating anorexic patients with an early age of onset. Some of these trials will be summarized in order to illustrate ethical and technical difficulties, as well as the strength of the evidence.

(a) The first controlled trial of family therapy

The ***first controlled trial of family therapy*** was begun at the Maudsley Hospital in the early 1980s on a series of 57 patients.[107] The principles of the family therapy have been incorporated in the Maudsley model, which has been adopted by researchers in the United States who have reached broadly similar conclusions.[108] One assumption of the Maudsley model of family therapy is that the family is ineffective in helping the patient eliminate her symptoms and might indeed contribute to their maintenance, because family life has become so organized around a potentially life-threatening problem. It is considered essential to correct the patient's starvation by assisting the parents to take control of their child's eating until her weight has returned to normal.

The first Maudsley trial had to overcome ethical objections to a randomized allocation bearing in mind the severity of the illness. Risks were minimized by first ensuring that the patients' weight loss had been corrected during an admission to an eating disorders unit. Family therapy itself commenced on discharge and aimed at the prevention of relapse. The outpatient family therapy was administered for 1 year and the patients were assessed on the completion of 1 year's treatment and 5 years later.

The 57 anorexic patients were subdivided into three groups according to recognized prognostic criteria. This showed that the benefits of family therapy were confined to one of the three subgroups, namely patients with the age of onset less than 19 years and a duration of illness less than 3 years.

Patients in each subgroup were randomly allocated to family therapy or the control treatment which consisted of a supportive and problem-centred individual therapy. Most important in assessing progress was body weight. The Morgan-Russell scales were utilized to obtain categories of general outcome and measures of adjustment along five dimensions: nutritional status, menstrual function, mental state, psychosexual adjustment and socio-economic status, as well as their mean, the 'average outcome score'.

(i) Group 1: early age of onset (short history)

A tendency for the patients to lose weight on discharge from hospital was reversed more readily in patients in receipt of family therapy. The superiority of family therapy was demonstrated by a higher weight at the end of 1 year's treatment (Fig. 4.10.1.2). Family therapy was also demonstrated to be more beneficial on the Morgan-Russell scales in terms of more good outcome categories and a higher average outcome score.

(ii) Group 2: patients with late age of onset

In this group the effect of the two therapies was reversed (Fig. 4.10.1.3). Individual therapy appeared to result in a greater weight gain than family therapy but significant only at a 6-months follow-up.

It was concluded that family therapy was more effective than individual therapy in patients whose illness began before the age of 19 years and had lasted less than 3 years.

(b) Beneficial components of family therapy in adolescents: the second Maudsley RCT[109]

The finding that family therapy is effective in younger patients with anorexia nervosa has led to a search for the effective components of this therapy. Many family therapists consider that it is important to understand the way the family functions. Others, however, prefer a symptom oriented approach with an emphasis on helping the parents to manage their child's problem.[43] The relative importance of these two components of family therapy has been investigated by comparing conjoint family therapy with separated family therapy.[109] In conjoint family therapy (CFT) the whole family is seen together in the treatment sessions. In separated family therapy (SFT) the parents are seen together and the same therapist also sees

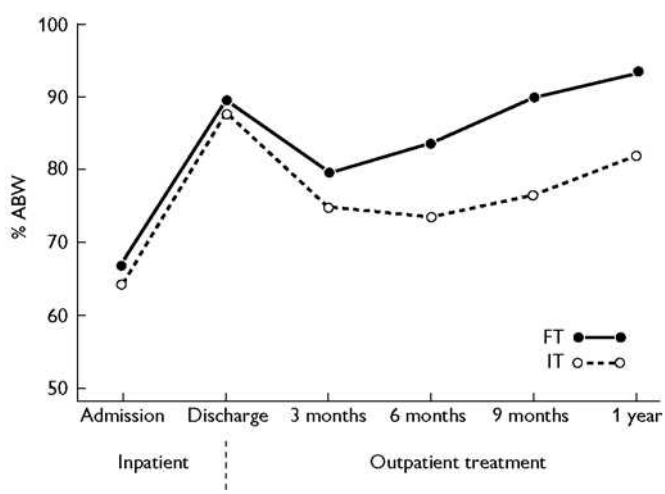


Fig. 4.10.1.2 Group of patients whose illness had an early onset and was of short duration. FT, family therapy; IT, individual therapy; ABW, average body weight.

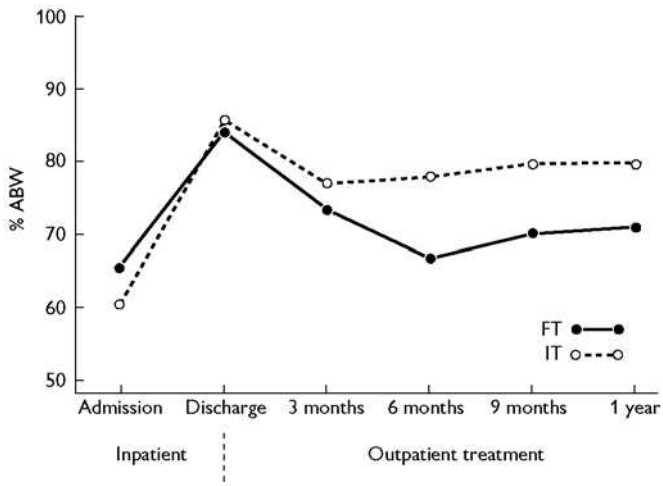


Fig. 4.10.1.3 Group of patients with relatively late onset of anorexia nervosa. FT, family therapy; IT, individual therapy; ABW, average body weight.

the patient separately for therapy and support. In CFT it is possible to observe and interpret family transactions, whereas this is not possible in SFT. The two treatments share the therapeutic advice for the parents to sustain a united approach aimed at improving their child's eating pattern.

The RCT was between CFT and SFT, given for 1 year on an outpatient basis. Forty adolescent patients were enrolled, most of them severely underweight (mean BMI 15.4).

In the course of the year the benefits of CFT and SFT were similar with considerable gains in weight (mean 8.2 kg) and marked improvements in the Morgan/Russell average outcome scores. Significant differences in the patients' responses to CFT and SFT were only obtained when treatment outcome was analysed separately in patients of high and low expressed emotion (EE) families. The patients within the high EE group allocated to SFT reached a significantly higher weight at follow-up (99.9 per cent ABW) than those in the CFT group (85.8 per cent ABW).

(c) RCT undertaken in Stanford, USA[108]

This trial was also on adolescent patients, this time comparing two courses of family therapy: one lasting for 6 months (10 sessions) and the other for 1 year (20 sessions). The aim of the study was to ascertain whether the duration of the family therapy exerted a critical effect on outcome.

The therapy was given according to the Maudsley Model, but the US study had a number of advantages, including a larger patient population (86 adolescents) and hence a greater statistical power. The therapy was given by therapists trained on a manual-based form of family therapy.

In comparison with the 2000 Maudsley Study[109] the patients were not so ill, (mean BMI 17.1). The principal measures of outcome were body weight and the eating disorder examination (EDE). There were no statistically significant differences between the two treatments.

The authors concluded that the short-term course of family was as effective as the long-term one for adolescents with a short duration anorexia nervosa. However, patients with more severe obsessive compulsive thinking, or coming from non-intact families, benefited from the longer treatment.

Strictly speaking, this RCT as such did not provide evidence of benefit from family therapy as there was no alternative form of treatment as a control. Nevertheless the outcomes at the end of 1 year's treatment were highly satisfactory with improvements not reported previously in natural outcome follow-up studies.

(d) Enduring benefits of family therapy

(i) 5-year follow-up with the first Maudsley RCT[110]

Patients who took part in the original Maudsley RCT of family therapy were followed up 5 years after the end of treatment in order to look for evidence of long-term benefit. Follow-up information was obtained on all 57 patients.[110] Apart from three deaths, there was a further overall improvement mainly attributable to the natural outcome of anorexia nervosa.

Within the subgroup of patients with early onset/short history, the mean weight achieved was 103 per cent ABW in patients who had received family therapy. On the Morgan/Russell scale, significantly greater improvement after family therapy was still discernable with a higher proportion of patients achieving a 'good' outcome and a higher average outcome score.

(ii) 5-year follow-up with the Maudsley adolescent RCT[112]

Of the original 40 patients 38 agreed to be reassessed. Compared with the end of treatment there were more patients in the good outcome category and fewer in the poor outcome category, and a further increase in weight. There were no deaths. The main purpose of this study had been to compare CFT and SFT. There was a significant difference between the two groups in the number of patients who resumed normal menstruation, higher in the SFT than in the CFT group. The findings at 1 year in patients from families with raised levels of maternal criticism (high EE) were still evident at 5 years. This group had done less well at the end of treatment in terms of weight and resumption of menstruation if they had been offered CFT.

(iii) 4-year follow-up with the US (Stanford) adolescent RCT[113]

Of the original 86 adolescent patients 71 were followed up. The long-term clinical outcome was good. Again there was little difference between patients treated with short-term or long-term family therapy. The BMI at follow-up for the short-term was 20.6, and the group as a whole showed good psychosocial functioning.

(e) Family therapy conclusions

There is now a strong body of evidence from RCTs that family therapy benefits younger patients with anorexia nervosa. The strength of the evidence in favour of family therapy has grown because of the increasing size of the patient populations tested: 21 in the 1987 Maudsley study, 40 in the 2000 Maudsley study, and 86 in the 2005 Stanford study.

1 From the search for specific components of the family therapy in the Maudsley studies, it can be concluded that the therapy should not be aimed at 'changing the family' but rather at helping them manage a sick family member.

2 Differences in the patients' responses to CFT and SFT were detected when there was a separate analysis for treatment outcome in patients in high and low expressed emotion (EE) families. From this it was concluded that SFT is the more appropriate treatment in families with high expressed emotion.

3 The Stanford study points towards the family therapy of shorter duration (6 months) being acceptable, except in the case of patients with obsessive compulsive thinking or in split families.

4 Finally, it is impressive that the quality of the clinical outcome increases in the long-term follow-ups. The patients in the 1987 Maudsley study showed the best outcome at the end of 5 years. This is not simply due to an expected good natural outcome, as the benefits of family therapy were still evident after 5 years.

Cognitive behavioural therapy

A theory for faulty cognitions maintaining the illness has been put forward by Fairburn *et al.*[114] The argument for examining the role of faulty cognitions in anorexia nervosa is inescapable. The original description of perceptual and conceptual disturbances in anorexia nervosa was put forward by Bruch in 1962.[5] It was appreciated that faulty attitudes to body size contributed in part to the patient's determination to reduce her food intake and lose weight. These observations led to the development of cognitive behavioural therapy (CBT) for anorexia nervosa. At first the evidence of its benefit relied on clinical impressions and case reports.[115] In recent years there have been valiant attempts to assess CBT in adult anorexic patients. It must be said, however, that these studies have met with limited success.

(a) CBT as a post-hospitalization treatment

The most fruitful RCT was conducted at the New York State Psychiatric Institute.[116] The design depended on the random allocation to CBT or the control treatment nutritional counselling (NC) of outpatients after discharge from an inpatient programme. Patients were eligible to enter the outpatient trial when they reached at least 90 per cent of ideal body weight. 33 who were eligible were randomly assigned to one or other treatment. The value of each treatment was assessed by its ability to prevent a relapse, defined as a fall in weight below a BMI of 17.5, or the development of medical or psychiatric complications.

Overall treatment failure was counted as the number of patients who had relapsed plus those who had dropped out from the treatment. The main finding is that the CBT group had a lower relapse and dropout rate and a better clinical outcome than the NC group.

(b) Comparison of CBT with a non-specific supportive clinical management (NSCM)

This study was conducted in Christchurch, New Zealand.[117] In this randomized controlled trial three therapies were compared:

1 CBT

2 Interpersonal psychotherapy

3 Non-specific supportive clinical management (NSCM).

The hypothesis was that the two specialized psychotherapies would be more effective than NSCM.

56 women were enrolled in the study. In fact the mean weights at baseline were not all that low (mean BMI 17.3). Twenty sessions were provided over a course of 20 weeks. The main assessments were a Global Assessment of Functioning (GAF) scale designed by the authors, body weight and the EDE.

The improvement in the group as a whole was perhaps disappointing with mean weight gains of 2.2 kg for CBT to 4.0 kg for NSCM. The hypothesis was not upheld and the specific therapies were less good than the non-specific supportive clinical management. The main lesson from this study is that good general clinical care combined with supportive psychotherapy can be at least as good as CBT.

(c) Multi-centred US study of CBT on its own or combined with medication[106]

An RCT was conducted in three well-established centres in the United States (White Plains New York, Minneapolis, and Stanford). In the original design there were three treatment groups:

1 Medication only

2 CBT only

3 Combined CBT and medication

Altogether 122 patients were randomized to one or other of these treatments. The patients were mainly adult. Their mean BMI (17.7–17.9) suggested the patients were not extremely thin. Intensive treatment was offered for up to 1 year. The patients were withdrawn from the study if they were 'treatment failures'. There were only low rates of completion: 27 per cent of the patients allocated medication, 43 per cent in the CBT group and 41 per cent in the combination group.

The methodological problems prevented an evaluation of the relative effectiveness of the treatments. The overall conclusions were somewhat pessimistic:

> 'It appears premature to conduct randomized controlled trials for *adults with anorexia nervosa* until the reasons for poor acceptance and high dropout rates . . . have been identified, and methods to remedy these serious problems have been devised'.

In fact the investigators were unlucky with the high rate of non-compliance; other studies offering psychotherapy to adult anorexic patients did not encounter such high dropout rates.[118]

The specialist eating disorder service

There is evidence that the treatment of anorexia nervosa is superior in specialized eating disorder units (EDUs) to that of general psychiatric units (adult or adolescent) or general medical units. The EDU comprises a team of professionals with training in the treatment of patients with anorexia nervosa and in maintaining the particular therapeutic ethos that is required. An EDU used to provide mainly inpatient care, but in recent years this has been extended to day and outpatient care. The clinical staff consist of a wide range of personnel including a psychiatric leadership and a trained nursing staff. The psychotherapeutic skills may vary but should include family therapy. The ethos of the EDU needs to combine a mix of therapeutic empathy for the patient with the ability to persuade her to return to a normal weight and eating habits. It is essential that the EDU nurse should be able to combine these skills.

This ethos was established in EDUs set-up in London in the 1970s and 1980s, particularly at St George's Hospital, the Royal Free Hospital and the Maudsley Hospital. It is from there that evidence has arisen for the effectiveness of a specialized eating disorder service.

An ambitious study from St George's Hospital, London, initially aimed at evaluating the advantages of three treatment settings.[96] Fortunately the study incorporated a control group and the

comparison with the active treatment groups yielded the most important results. The four groups were as follows:

1 Inpatient treatment: 4 months on average followed by outpatient psychotherapy.

2 Outpatient treatment combining individual and family therapy.

3 Outpatient group therapy.

4 The control group provided no treatment at the EDU after a 'one-off' evaluation.

90 patients were randomly allocated to one or other of the four options. The study began in the mid-1980s when uncertainty in the subject still warranted from an ethical point of view a random allocation to four groups, including one in which no treatment was provided but the patients returned to the care of their general practitioners. There were methodological difficulties. The patients sometimes dropped out of the treatment when it was not what they wished: out of 30 patients offered inpatient treatment, only 18 accepted it.

At 1 year follow-up there were few clear-cut differences between the patients in the three treatment groups: the inpatients' weights were similar to the outpatients. The clearest finding was that patients allocated to any one of the three treatment groups fared much better than those allocated to the one-off evaluation session only.

The main lesson from this valiant study is that care within an ED service, whether inpatient or outpatient, and irrespective of the specific treatment modality, is superior to treatment received outside an ED service.

(a) Day care and community treatment

There is a paucity of controlled trials of day programmes. An exception is the comparison of inpatient and day treatment carried out in Edinburgh.[119] A traditional inpatient programme, aimed strictly at weight gain, was compared with a more permissive day programme stressing open communication and patient autonomy. The day programme consisted of intensive psychological treatments and was available on 4 days each week. 32 patients who would have merited admission to hospital were randomly allocated to the day programme or the inpatient programme. Although statistically significant differences were not found between the two methods of treatment, the author reported interesting trends in improved outcome using the Morgan/Russell scales. The only advantage conferred by inpatient treatment was a slightly greater weight gain. In contrast, the day programme was more popular with patients who preferred to have a greater say in their rate of weight gain. There was also a greater return of personal autonomy.

(b) Inpatient treatment

Inpatient treatment in recent years has attracted a bad press. It has been argued that there are few RCTs, which have thrown any light on the benefits of inpatient treatments. This is a blinkered approach because RCTs addressed to some of the questions, are either inappropriate or ethically impossible to carry-out. An example of a largely inappropriate question from a clinical point of view is the comparison of inpatient and outpatient treatment in anorexia nervosa because the indications and the patient's suitability for each are very different. It is justified to state this categorically because there now exists solid clinical experience to show what can be achieved by inpatient treatment. The limitations of inpatient treatment are that the benefits may last only a number of months when there is a variable likelihood of relapse. A useful research question would be to identify patients who are most likely to relapse after inpatient treatment. The firm statement made above is warranted because of the valiant study by Crisp and his colleagues[120] previously discussed. Today this study would probably not be possible from an ethical point of view.

The most obvious benefit from inpatient treatment in an EDU is that this is the surest and safest method of improving the patient's nutrition and correcting her weight loss, thereby reversing physical complications, and sometimes saving her life. In various cohort studies the weight gain has varied around 12 kg in 12 weeks.[120–122] There are no comparable rapid improvements with outpatient treatments, not even family therapy in adolescents.

There have been a small number of RCTs in anorexia nervosa in addition to the Crisp *et al* study, and they have addressed narrower questions. For example, a comparison of two different inpatient treatment programmes has yielded the helpful finding that a strict operant conditioning programme offers no advantages over a 'lenient' programme.[123] The value of a specialized eating disorder unit in inducing weight gain was also demonstrated in a cohort of patients admitted to the Maudsley Hospital in 1990. They had previously been admitted to a general psychiatric or medical unit, and information on their previous weights was obtained. In the case of the Maudsley admissions the mean weight rose from 65.8 per cent average body weight (abw) to 90.4 per cent abw; in comparison the admissions elsewhere only led to a mean weight gain from 64.4 to 75 per cent abw.[122]

(c) Compulsory treatments

A study of the use of compulsory treatment in patients admitted to the Eating Disorders Unit at the Maudsley Hospital, comprised 81 patients or16 per cent of admissions. Section 3 of the Mental Health Act, valid for up to 6 months, was the most frequently applied section.[121] The compulsorily admitted patients were compared with a group of voluntary patients. The need for a compulsory admission was found to have two dimensions—the presence of a severe illness and a rejection of treatment. The compulsory patients gained at least as much weight as the voluntary patients but required a longer admission for them to return to a near-normal weight. The compulsory patients represented a selected group by virtue of a more entrenched reluctance to accept treatment. Accordingly it was predicted that in the long term they would fare less well than voluntary patients. The mortality rate of these patients was determined by the National Register, which provided the data at a mean of 5.7 years after the index admission. Ten out of 79 detained patients had died in comparison with two out of 78 voluntary patients, a statistically significant difference. The deaths among the compulsory patients were all due to anorexia nervosa or one of its nutritional complications. Thus the mortality rate among compulsory patients was extremely high at 2.17 per cent per annum, presumably skewed because of the selection factors. Compulsory treatment is an obvious example of the inappropriateness of an RCT.

Treatment: advice on management

It is not possible for the clinician to confine his treatment of anorexia nervosa to those methods that have been subjected to

randomized control trials. This does not mean that his treatment is ineffective or unsupported by evidence. The evidence has to be derived from other kinds of studies. They should not be dismissed as mere clinical experience if they can be supported by evidence-based clinical or experimental observations.

The main obstacle to treatment

The avoidance of treatment by the patient is part and parcel of anorexia nervosa and accentuated by her capacity for denial (see p. 783 of this chapter). There have been attempts to predict the likely level of the patient's compliance with treatment. For example, a 'transtheoretical model of change' is aimed at improving the patient's motivation by overcoming ambivalence while at the same time avoiding confrontation.[124] Different stages in the patient's approach to treatment have been recognized: precontemplation, contemplation, preparation, action, maintenance, and relapse prevention. Motivational enhancement therapy is at an early stage of development with only preliminary information on its impact on treatment outcome.

At our present level of knowledge it is simplest to ascertain the patient's attitude to treatment through the mental state examination. At the initial interview she should be asked what weight she would be willing to reach. At this stage it is best to refrain from challenging the patient's weight threshold, even though it will be well below a reasonable therapeutic goal. Having ascertained the limited degree of compliance, the clinician needs to develop a strategy to improve it gradually as treatment proceeds. It is poor clinical practice to place all the onus on the patient herself for accepting or rejecting a package of treatment at the first interview, or even at a later stage. 'Engaging' the patient in treatment is an essential part of most psychotherapeutic methods including CBT.[115] Another tactic for improving the patient's co-operation is to enlist the help of close members of the family. Young patients are likely in any case to require a family treatment.

Inpatient treatment

Although inpatient treatment is less often employed nowadays it will be described first because it is most important in patients who present as emergencies with an urgent need to preserve life and correct serious physical and psychiatric complications.

(a) Indications for admission

- It is tempting to specify a BMI threshold for admission, but it should be remembered that the BMI can only provide a very rough guide with many exceptions. If the BMI is below 14 admission should always be considered, all the more if there has been rapid weight loss during recent weeks. Indeed a BMI of 16 may cause anxiety if it represents a drop in weight from a BMI of 21 in the course of 2 or 3 months. Close attention should be given to the physical manifestations of malnutrition, especially hypoglycaemia and electrolyte disturbances.
- **The BMI may be totally unreliable in patients who have developed oedema as the result of malnutrition, causing a vicarious weight gain, which may deceive the patient and even the clinician into a false sense of security.**
- The BMI is also less reliable in children and adolescents, especially if their growth in height has been retarded thus distorting the calculation for the BMI towards a falsely reassuring value.
- Dangerous or disabling physical complications:
- Emergencies may result from low blood glucose levels, hypokalaemia or hyponatraemia. The electrolyte disturbances are most likely to occur in the presence of self-induced vomiting or laxative abuse. Hyponatraemia may be induced by polydipsia. Signs of proximal myopathy also indicate a need for admission, as do the occurrence of anaemia, low platelet counts (sometimes accompanied by purpura) and impaired liver function tests.
- Severe depression or obsessional behaviour may also arise in part as complications of malnutrition. Suicidal ideas may require admission, but persistent depression of lesser severity may also be an indication, because it renders the patient less amenable to outpatient psychological treatments, as does severe obsessional behaviour affecting the patient's eating pattern.
- Intractable self-induced vomiting or laxative abuse or determined fasting may require a greater supervision than is possible in an outpatient setting. These behaviours are less disconcerting in patients with bulimia and at a reasonable weight, but a severe weight loss makes it probable that vomiting will result in electrolyte disturbances or other physical complications.
- The decision to admit is often determined by the patient not engaging in the recommended outpatient psychological treatment when the malnutrition persists several months to 1 year.
- Young patients with a retarded puberty (premenarchal anorexia nervosa) should elicit an urgent correction of malnutrition, especially if there is evidence of retarded growth and short stature. There is a narrow window of opportunity for regaining stature, which may depend partly on bone age but it is safest to assume that after the age of 16 the probability of resuming growth becomes increasingly limited, especially in girls.

The great advantage of treating a patient in a specialized eating unit is the certainty that considerable benefit will accrue, including a substantial gain in weight, so long as the patient can be persuaded to remain in hospital.[120]

(b) Nursing and dietary care

Inpatient treatment will include a wide range of psychotherapeutic interventions (individual, group, and family) as well as occupational therapy and an educational programme. But the sheet anchor of successful treatment is a well-trained nursing staff working as a team. The role of the medical staff is one of maintaining a high level of expectation that the patient's weight will be restored to a normal (or healthy) level. It is necessary to give the nursing staff moral support so that they can develop their confidence and skills.

There are two main components to the nurses' treatment: their psychotherapeutic input and their supervisory role. The latter should never be draconian. The nursing team establish a relationship of trust with the patient and get to know her personal needs and concerns. The nurses should also be acutely aware of the anorexic patient's tendency to avoid food and exercise excessively. The treatment programme should stress the supportive aspects of the nurse's relationship with her patient rather than the undoubted need for careful supervision. The nurse will come to rely on the daily weight record to monitor the success of her treatment as the weight chart should show a smoothly rising curve. All meals are taken in the ward: thus the anorexic patients constitute a group in which peer interactions take place. The meal is taken in the presence

of one or more nurses also seated at the table. The patient learns that she is expected to consume all the food placed before her.

A detailed protocol for the refeeding programme has been produced by Andersen and his group.[120] It is not only the patient who tends to underestimate the food requirements to restore her weight to normal. Metabolic studies have demonstrated that for each kilogram of weight gain a surplus calorie balance of 7500 cal is needed.[111] It is prudent to begin with a modest calorie intake of 1200 to 1500 cal daily during the first 7 days in order to avoid the complications of hypophosphataemia and acute gastric dilation. Thereafter the caloric intake is gradually increased and may rise to 4000 cal. daily. The best diet is that consisting of a wide range of foods including carbohydrate and fat-containing foods. Concentrated foods (e.g. Build-up, Complan, Scandishake, Fortisip, Ensure Plus) may be used to achieve a high caloric intake. The aim is to achieve a positive energy balance of at least 1500 cal daily, leading to a weekly weight gain of 1 to 1.5 kg.

(c) Assessment of progress

Weighing should be a standardized daily procedure before breakfast after the patient has emptied her bladder and while wearing light night clothes. A paradoxical psychological improvement, with a diminution in concern with body size and shape, occurs with weight gain. The improvement is partly through the correction of malnutrition and partly the result of the 'exposure treatment' whereby the patient gradually accepts a higher body weight.

(d) Medication

Exceptionally the patient's tension and depression do not improve and there is a continued resistance to food. It may then be helpful to prescribe moderate doses of olanzapine (not more than 10 mg daily), carefully avoiding a fall in blood pressure, which is a risk in the emaciated patient. In the case of persistent depression, treatment with an antidepressant may be indicated. However, antidepressants are often ineffective in the presence of malnutrition, and by themselves do not assist the patient to gain weight.

(e) General measures

Inevitably the patient will find it irksome to forego home visits during the early period of weight gain. Therefore interesting and therapeutic activities should be provided through group meetings, occupational therapy, and social interactions. Visiting is generally encouraged unless the visiting parents are subjected to emotional appeals to take her home. They may then be asked to postpone their visits or reduce their duration. The monotony can be relieved when the patient's weight gain is on course by providing accompanied outings avoiding mealtimes.

The aim is to restore body weight to a healthy level within 8 to 10 weeks; a further period (usually 2 weeks) in hospital is needed to allow the patient to test her ability to maintain her weight by eating in a general dining room or going on home leave for two or more days at a time. The aim is to effect a smooth transition to day care or outpatient treatment.

Compulsory treatment in anorexia nervosa

The management of patients reluctant to accept essential therapeutic goals requires that they should be gradually engaged in a genuine alliance. However, there remain a minority of patients with whom this strategy fails and whose health and life become endangered. For them, compulsory treatment should be considered.[121] A compulsory admission to hospital is indicated not only when patients threaten suicide or suffer from a life-threatening malnutrition, but also when they fail to respond to simpler measures such as outpatient treatment or day care, or in the event of avoiding treatment altogether. Ill health persisting over the course of several months or the development of serious physical complications (e.g. water and electrolyte imbalance, hypoglycaemia, or myopathy) should also elicit compulsory admission if the patient cannot be persuaded to accept inpatient treatment voluntarily.

The evidence points to the usefulness of a compulsory admission in appropriate circumstances in so far as the patients respond well in the short term. Nevertheless, a patient who has required a compulsory admission carries a higher risk, so that it is essential to safeguard her through a long follow-up.

Day and community treatment

The Edinburgh trial of a day patient programme has already been discussed. At the Toronto Hospital a day hospital programme for eating disorders has been established since 1985.[125] and is now offered on 4 days a week.

The goals are as follows:

1 A normalization of disturbed eating behaviour and weight gain.

2 The identification of psychological and familial processes that serve to perpetuate the eating disorder.

Two meals and a snack are provided during the treatment hours. The staff take turns to supervise the patients during meal times. The psychological treatment consists of intensive group therapy addressing disturbed behaviours around eating and weight, and more general conflicts.

The clinical advantages of day treatment are a reduction in the dependence of patients who need to be able to function outside the hospital. The group treatment provides an atmosphere of mutual support while permitting interventions through group pressures. A wide range of patients can be admitted but those with medical or suicidal risks will elicit inpatient care instead. When patients succeed in reducing their disturbed eating behaviours they may 'act out' by self-harm. The clinical staff may find their skills severely taxed by the continuous staff/patient interaction.

A home oriented service extended to outpatient and day care, has been devised by Robinson (2006) for an area of North London (1.2 million). A wide range of treatments were devised including family interventions and carer support. The educational needs of members of the multi-disciplinary team are well described as are administrative and financial issues. Robinson's book includes an audit on the use of hospital and hostel beds requiring only 4–5 total beds per million population per year.

Outpatient psychotherapies

In the event of a patient's weight loss being less than 20 per cent of her healthy weight, it may be possible to obtain a therapeutic response by outpatient treatment, including attendance at non-specialist general psychiatric clinics or child/adolescent psychiatric services. The feasibility of this approach will depend on the availability of appropriate psychotherapeutic resources. The clinician should guard himself against imaginary conflicts between a psychotherapeutic approach and recording the weight of the patient. It is never justified to accept apparent compliance with psychotherapy if the patient's weight continues to decline.

(a) Family treatments

(i) Conjoint family therapy

The frequency of the sessions is determined by clinical need: it averaged 10.5 over the course of 1 year in the Maudsley trial. There are three stages to the therapy:

1 In the first phase the parents are urged to identify their future joint attitude to the feeding pattern that should be adopted by their daughter. With a younger patient the therapist assumes that the parents would initially need to take control of their child's eating.

2 During the second phase the parents are urged to be present together at each meal so that they can reinforce each other's efforts in the practical task of ensuring an improved food intake and a steady weight gain.

3 When the patient's weight is under control, responsibility for continued weight gain is handed back to her. Discussions can then commence on more normal family concerns. With an adolescent patient the main focus is on achieving increased autonomy. The eventual aim is to establish healthy relationships with her parents without the eating disorder as a medium for communication.

(ii) Separated family therapy

As in conjoint family therapy the parents are given direct advice on how to manage their daughter's eating problem. The patient herself is provided with individual psychotherapy. The therapist provides counselling about abnormal attitudes to weight and emphasizes the weight issue until steady progress has been made. This method is often preferred by the patient and her parents, largely because it avoids confrontation. It is also easier to gain access to the patient's fears and conflicts.

(b) CBT

CBT has much in common with other methods of treatment including the refeeding programme described under inpatient treatments. It relies on building a positive therapeutic alliance between therapist and patient.

The patient's weight and food intake is monitored at each session and she is told her weight. She is encouraged to think of food as medication and to follow a meal plan. The patient is educated in the disturbances of bodily and psychological function consequent on the state of starvation. The content of the therapy may be divided into two 'tracks'. The first track includes an examination of the behaviours adopted by the patient in order to reduce her weight. The second track is more concerned with psychological themes such as self-esteem, perfectionism, interpersonal functioning, and family conflicts. By asking the patient to give her reasons for specific behaviours, the therapist discovers faulty beliefs and assumptions on her part. For example, the 'anorexic wish' is the patient's wish to recover from her disorder without gaining weight. She is gently persuaded that this is an impossible aim because her real psychological difficulties will remain inaccessible so long as her experiences are clouded by starvation and dieting. The patient also expresses a fear of 'losing control' meaning that she will run the risk of overeating and become fat. It is explained to the patient that her rigid 'control' overeating deprives her completely of choice, and that far from being in control the reality is the converse. It is also useful for the therapist to analyse the pros and cons of maintaining the disorder of anorexia nervosa. The patient often feels more uncomfortable at confronting the hidden rewards of remaining thin.

Having ascertained the particular meanings of attitudes and behaviours for the patient, she is helped to find more adaptive ways of achieving healthier goals, including more relaxed normal eating and weight gain.

(c) Dynamic psychotherapy: therapeutic model

(i) Crisp's model of anorexia nervosa as a flight from growth

Crisp has explained how his programme of treatment is based on his model of anorexia nervosa as a refuge from puberty, which the patient has found overwhelming. The youngster reverses her pubertal development by limiting her intake of food.[126] His treatment initially involves an intensive inpatient programme lasting 10 to 12 weeks followed by outpatient treatment for up to 6 years. The advantage of this extensive programme is that it enables the patient to accept gradually an increase in weight while facing up to the feelings of panic and helplessness that originally led her to arrest her puberty through self-starvation. This interpretation is presented to the patient and to her family so that they come to see the psychotherapy as a way of solving the problems.

The model is a useful one even within a more limited outpatient psychotherapeutic setting. Some patients will readily identify their distress when overwhelmed by powerful sexual feelings or when confronted with personal and social responsibilities perceived as the result of growing up. The aims of the psychotherapy are to support the patient while she is beginning to abandon the psychobiological regression of anorexia nervosa. In addition she is encouraged to broaden her perception of herself in ways that are no longer wholly dependent on her physical appearance but include an improved sense of competence and self-esteem. She is helped to tackle personal and social problems from which she had previously escaped so that she can address her own and her parents' concerns about sexuality.

Ethical and medico-legal issues

In the United Kingdom the Mental Health Act Commission[127] has clarified many of the doubts in the minds of clinicians and social workers called upon to consider a compulsory admission under the Mental Health Act 1983. It recognized that anorexia nervosa is a mental disorder within the meaning of the Act and that in some patients their ability to consent may be compromised by fears of obesity or denial of the consequences of their actions. The Mental Health Act Commission concluded that when the patient's health is seriously threatened by food refusal she may be detained in hospital so as to treat the self-imposed starvation. The Commission went as far as to state that nasogastric feeding can be a medical process forming an integral part of the treatment for anorexia nervosa, notwithstanding that nasogastric feeding is seldom required.

Prevention

Preventive measures have been aimed at eating disorders generally rather than just anorexia nervosa. The main approach has been school-based intervention programmes educating adolescent girls into the risks of dieting and other methods of weight control. The commonest outcome has been an increased knowledge about

eating disorders without a change in the behaviours, such as dieting, likely to cause them. One controlled study, six weekly sessions were provided by teachers on a wide range of topics covering attitudes and behaviours relevant to eating disorders including a 'non-dieting approach'. Positive changes were detected at a 6-month follow-up but they were modest in size and poorly sustained.[128]

Currently there are active campaigns by the Academy of Eating Disorders and Beat Eating Disorders (UK) to encourage the fashion industry not to employ thin models. Unilever has adopted a policy not to employ models with a BMI less than 18.5. John Lewis requires models to be not less than size 12.

Further information

Robinson, P.H. (2006). *Community treatment of eating disorders*. John Wiley, Chichester.

Russell, G.F.M. (1993). Social psychiatry of eating disorders. In *Principles of social psychiatry* (eds, D. Bhugra and J. Leff). pp. 273–97. Blackwell Science, Oxford.

References

1. Gull, W.W. (1874). Anorexia nervosa (apepsia hysterica, anorexia hysterica). *Transactions of the Clinical Society of London*, **7**, 22–8.
2. Lasègue, C. (1873). De l'anorexie hystérique. *Archives Générales de Médicine*, **21**, 385–403.
3. Brumberg, J.J. (1988). *Fasting girls: the emergence of anorexia nervosa as a modern disease*. Harvard University Press, Cambridge, MA.
4. Bell, R.M. (1985). *Holy anorexia*. University of Chicago Press, Chicago and London.
5. Bruch, H. (1962). Perceptual and conceptual disturbances in anorexia nervosa. *Psychosomatic Medicine*, **24**, 187–94.
6. Russell, G.F.M. (1970). Anorexia nervosa: its identity as an illness and its treatment. In *Modern trends in psychological medicine* (ed. J. Harding Price), pp. 131–64. Butterworths, Norwich.
7. Garner, D.M. and Garfinkel, P.E. (1979). The eating attitudes test: an index of the symptoms of anorexia nervosa. *Psychological Medicine*, **9**, 273–9.
8. Johnson-Sabine, E., Wood, K., Patton, G., *et al.* (1988). Abnormal eating attitudes in London schoolgirls—a prospective epidemiological study: factors associated with abnormal response on screening questionnaires. *Psychological Medicine*, **18**, 615–22.
9. Hoek, H.W. (1991). The incidence and prevalence of anorexia nervosa and bulimia nervosa in primary care. *Psychological Medicine*, **21**, 455–60.
10. van Son, G.E., van Hoeken, D., Bartelds, A.I., *et al.* (2006). Time trends in the incidence of eating disorders: a primary care study in the Netherlands. *The International Journal of Eating Disorders*, **39**, 565–9.
11. Garner, D.M. and Garfinkel, P.E. (1980). Socio-cultural factors in the development of anorexia nervosa. *Psychological Medicine*, **10**, 647–56.
12. Szmukler, G.I., Eisler, I., Gillies, C., *et al.* (1985). The implications of anorexia nervosa in a ballet school. *Journal of Psychiatric Research*, **19**, 177–82.
13. Crisp, A.H., Palmer, R.R.L., *et al.* (1976). How common is anorexia nervosa? A prevalence study. *The British Journal of Psychiatry*, **128**, 549–54.
14. Råstam, M., Gillberg, C., and Garton, M. (1989). Anorexia nervosa in a Swedish urban region. A population based study. *The British Journal of Psychiatry*, **155**, 642–6.
15. Russell, G.F.M. (1993). Social psychiatry of eating disorders. In *Principles of social psychiatry* (eds. D. Bhugra and J. Leff), pp. 273–97. Blackwell Science, Oxford.
16. Milos, G., Spindler, A., Schnyder, U., *et al.* (2004). Incidence of severe anorexia nervosa in Switzerland: 40 years of development. *The International Journal of Eating Disorders*, **35**, 250–61.
17. Lucas, A.R., Beard, C.M., O'Fallen, W.M., *et al.* (1991). 50-year trends in the incidence of anorexia nervosa in Rochester, Minn: a population-based survey. *The American Journal of Psychiatry*, **148**, 917–22.
18. Fosson, A., Knibbs, J., Bryant-Waugh, R., *et al.* (1987). Early onset anorexia nervosa. *Archives of Disease in Childhood*, **62**, 114–18.
19. Favaro, A., Ferrara, S., and Santonastaso, P. (2003). The spectrum of eating disorders in young women: a prevalence study in a general population sample. *Psychosomatic Medicine*, **65**, 701–8.
20. Blake Woodside, D., Garfinkel, P.E., Lin, E., *et al.* (2001). Comparisons of men with full or partial eating disorders, and women with eating disorders in the community. *The American Journal of Psychiatry*, **158**, 570–4.
21. Morgan, H.G. and Russell, G.F.M. (1975). Value of family background and clinical features as predictors of long-term outcome in anorexia nervosa: four-year follow-up study of 41 patients. *Psychological Medicine*, **5**, 355–71.
22. Russell, G.F.M. (1979). Bulimia nervosa: an ominous variant of anorexia nervosa. *Psychological Medicine*, **9**, 429–48.
23. Russell, G.F.M. (1995). Anorexia nervosa through time. In *Handbook of eating disorders: theory, treatment and research* (eds. G. Szmukler, C. Dare, and J. Treasure), pp. 5–17. Wiley, Chichester.
24. Currin, L., Schmidt, U., Treasure, J., *et al.* (2005). Time trends in eating disorder incidence. *The British Journal of Psychiatry*, **186**, 132–5.
25. Campbell, P.G. (1995). What would a causal explanation of the eating disorders look like? In *Handbook of eating disorders: theory, treatment and research* (eds. G. Szmukler, C. Dare, and J. Treasure), pp. 49–64. Wiley, Chichester.
26. Garfinkel, P.E. and Garner, D.M. (1982). *Anorexia nervosa: a multidimensional perspective*. Brunner-Mazel, New York.
27. Birnbaum, K. (1923). *Der Aufbau der Psychose*, pp. 6–7. Springer, Berlin.
28. Patton, G.C., Johnson-Sabine, E., Wood, K., *et al.* (1990). Abnormal eating attitudes in London schoolgirls-a prospective epidemiological study: outcome at twelve month follow-up. *Psychological Medicine*, **20**, 383–94.
29. Crisp, A.H. (1970). Premorbid factors in adult disorders of weight, with particular reference to primary anorexia nervosa (weight phobia). A literature review. *Journal of Psychosomatic Research*, **14**, 1–22.
30. Prince, R. (1985). The concept of culture-bound syndromes: anorexia nervosa and brain-fag. *Transcultural Psychiatric Research Review*, **22**, 117–21.
31. DiNicola, V.F. (1990). Anorexia multiforme: self-starvation in historical and cultural context. *Transcultural Psychiatric Research Review*, **27**, 165–96, 245–86.
32. Bryant-Waugh, R. and Lask, B. (1991). Anorexia nervosa in a group of Asian children living in Britain. *The British Journal of Psychiatry*, **158**, 229–33.
33. Mynors-Wallis, L., Treasure, J., and Chee, D. (1992). Life events and anorexia nervosa: differences between early and late onset cases. *The International Journal of Eating Disorders*, **4**, 369–75.
34. Beumont, P.J.V., Abraham, S.F., and Simson, K.G. (1981). The psychosexual histories of adolescent girls and young women with anorexia nervosa. *Psychological Medicine*, **11**, 131–40.
35. Tiller, J. and Treasure, J. (1998). Eating disorders precipitated by pregnancy. *European Eating Disorders Review*, **6**, 178–87.
36. Sloan, G. and Leichner, P. (1986). Is there a relationship between sexual abuse or incest and eating disorders? *Canadian Journal of Psychiatry*, **31**, 656–60.
37. Palmer, R.L., Oppenheimer, R., Dignon, A., *et al.* (1990). Childhood sexual experiences with adults reported by women with eating disorders: an extended series. *The British Journal of Psychiatry*, **156**, 699–703.

38. McClelland, L., Mynors-Wallis, L., Fahy, T., *et al.* (1991). Sexual abuse, disordered personality and eating disorders. *The British Journal of Psychiatry*, **158**(Suppl. 10), 63–8.
39. Wonderlich, S.A., Brewerton, T.D., Jocic, Z., *et al.* (1997). Relationship of childhood sexual abuse and eating disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**, 1107–15.
40. Minuchin, S., Baker, L., Rosman, B.L., *et al.* (1975). A conceptual model of psychosomatic illness in children: family organization and family therapy. *Archives of General Psychiatry*, **32**, 1031–8.
41. Selvini Palazzoli, M. (1974). *Self-starvation. From the intrapsychic to the transpersonal approach to anorexia nervosa* (trans. A. Pomerans), pp. 202–16. Human Context Books, London.
42. Bruch, H. (1978). *The golden cage: the enigma of anorexia nervosa*. Open Books, London.
43. Dare, C. and Eisler, I. (1997). Family therapy for anorexia nervosa. In *Handbook of treatment for eating disorders* (2nd edn) (eds. D.M. Garner and P.E. Garfinkel), pp. 307–24. Guilford Press, New York.
44. Janet, P. (1903). L'obsession de la honte du corps. In *Les obsessions et la psychasthénie*, Vol. 1, Section 5. Germer Gaillière, Paris.
45. Wonderlich, S.A., Swift, W.H., Slotnick, H.B., *et al.* (1990). DSM-III-R personality disorders in eating disorder subtypes. *The International Journal of Eating Disorders*, **9**, 607–16.
46. Casper, R.C. (1990). Personality features of women with good outcome from restricting anorexia nervosa. *Psychosomatic Medicine*, **52**, 156–70.
47. Simmonds, K. (1916). Über kachexie hypophysaren ursprungs. *Deutsche Medizinische. Wochenschrift*, **42**, 190–1.
48. Russell, G.F.M. (1977). The present status of anorexia nervosa. *Psychological Medicine*, **7**, 363–7.
49. Anand, B.K. and Brobeck, J.R. (1951). Localization of a 'feeding center' in the hypothalamus of the rat. *Proceedings of the Society for Experimental Biology and Medicine*, **77**, 323–4.
50. Schweiger, U., Tuschl, R.J., Laessle, R.G., *et al.* (1989). Consequences of dieting and exercise on menstrual function in normal weight young women. In *The menstrual cycle and its disorders* (eds. K.M. Pirke, W. Wuttke, and U. Schweiger), pp. 142–9. Springer Verlag, Berlin.
51. DeVile, C.H., Sufraz, R., Lask, B., *et al.* (1995). Occult intracranial tumours masquerading as early onset anorexia nervosa. *British Medical Journal*, **311**, 1359–60.
52. Ploog, D.W. and Pirke, K.M. (1987). Psychobiology of anorexia nervosa. *Psychological Medicine*, **17**, 843–59.
53. Lask, B., Gordon, I., Christie, D., *et al.* (2005). Functional neuroimaging in early onset anorexia nervosa. *The International Journal of Eating Disorders*, **37**, S49–51.
54. Strober, M., Lampert, C., Morrell, W., *et al.* (1990). A controlled family study of anorexia nervosa: evidence of familial aggregation and lack of shared transmission with affective disorders. *The International Journal of Eating Disorders*, **9**, 239–53.
55. Bulik, C.M. (2004). Role of genetics in anorexia nervosa, bulimia nervosa and binge eating disorder. In *Clinical handbook of eating disorders* (ed. T.D. Brewerton), pp. 165–82. Marcel Dekker, New York.
56. Holland, A.J., Hall, A., Murray, R., *et al.* (1984). Anorexia nervosa: a study of 34 twin pairs and one set of triplets. *The British Journal of Psychiatry*, **145**, 414–19.
57. Holland, A.J., Sicotte, N., and Treasure, J. (1988). Anorexia nervosa: evidence for a genetic basis. *Journal of Psychosomatic Research*, **32**, 561–71.
58. Blundell, J.E. and Bill, A.J. (1991). Serotonin, eating disorders and the satiety cascade. In *Serotonin-related psychiatric syndromes: clinical and therapeutic links* (eds. G.B. Cassano and H.S. Asikal), pp. 125–9. Royal Society of Medicine Services, London.
59. Szmukler, G.I., Andrewes, D., Kingston, K., *et al.* (1992). Neuropsychological impairment in anorexia nervosa before and after refeeding. *Journal of Clinical and Experimental Neuropsychology*, **14**, 347–52.
60. Kaye, W.H., Gwirtsman, J.E., George, D.T., *et al.* (1988). CSF-5HIAA concentrations in anorexia nervosa: reduced values in underweight subjects normalize after weight gain. *Biological Psychiatry*, **23**, 102–5.
61. Kaye, W.H., Gwirtsman, H.E., George, D.T., *et al.* (1991). Altered serotonin activity in anorexia nervosa after long-term weight restoration: does elevated CSF-5HIAA correlate with rigid and obsessive behavior? *Archives of General Psychiatry*, **48**, 556–62.
62. O'Dwyer, A.-M., Lucey, J.V., and Russell, G.F.M. (1996). Serotonin activity in anorexia nervosa after long-term weight restoration: response to d-fenfluramine challenge. *Psychological Medicine*, **26**, 353–9.
63. Slade, P.D. and Russell, G.F.M. (1973). Awareness of body dimensions in anorexia nervosa: cross-sectional and longitudinal studies. *Psychological Medicine*, **3**, 188–99.
64. Hsu, L.K.G. and Sobkiewicz, T.A. (1989). Body image disturbance: time to abandon the concept for eating disorders? *The International Journal of Eating Disorders*, **10**, 15–30.
65. Hsu, L.K.G. and Lee, S. (1993). Is weight phobia always necessary for a diagnosis of anorexia nervosa? *The American Journal of Psychiatry*, **150**, 1466–71.
66. Blake Woodside, D. and Twose, R. (2004). Diagnostic issues in eating disorders: historical perspectives and thoughts for the future. In *Clinical handbook of eating disorders* (ed. T.D. Brewerton), pp. 1–19. Marcel Dekker, New York.
67. Vandereycken, W. (2007). Denial of illness in anorexia nervosa—a conceptual review: part I diagnostic significance and assessment. *European Eating Disorder Review*, **14**, 352–68.
68. Fichter, M.M. and Pirke, K.M. (1995). Starvation models and eating disorders. In *Handbook of eating disorders: theory, treatment and research* (eds. G. Szmukler, C. Dare, and J. Treasure), pp. 83–107. Wiley, Chichester.
69. Treasure, J.L., Gordon, P.A.L., King, E.A., *et al.* (1985). Cystic ovaries: a phase of anorexia nervosa. *Lancet*, **2**, 1379–82.
70. Treasure, J.L., Wheeler, M., King, E.A., *et al.* (1988). Weight gain and reproductive function: ultrasonographic and endocrine features in anorexia nervosa. *Clinical Endocrinology*, **29**, 607–16.
71. Namir, S., Melman, K.N., and Yager, J. (1986). Pregnancy in restrictor-type anorexia nervosa: a study of six women. *The International Journal of Eating Disorders*, **5**, 837–45.
72. Treasure, J.L. and Russell, G.F.M. (1988). Intrauterine growth and neonatal weight gain in babies of women with anorexia nervosa. *British Medical Journal*, **296**, 1038.
73. Van Wezel-Meijler, G. and Wit, J.M. (1989). The offspring of mothers with anorexia nervosa: a high-risk group for undernutrition and stunting? *European Journal of Pediatrics*, **49**, 130–5.
74. Treasure, J. and Szmukler, G. (1995). Medical complications of chronic anorexia nervosa. In *Handbook of eating disorders: theory, treatment and research* (eds. G. Szmukler, C. Dare, and J. Treasure), pp. 197–220. Wiley, Chichester.
75. Passmore, R. and Eastwood, M.A. (1986). *Human nutrition and dietetics*. Churchill Livingstone, Edinburgh.
76. Handler, C.E. and Pirkin, G.D. (1982). Anorexia nervosa and Wernicke's encephalopathy: an undiagnosed association. *Lancet*, **2**, 771–2.
77. Lask, B. and Bryant-Waugh, R. (1992). Early onset anorexia nervosa and related eating disorders. *Journal of Child Psychology and Psychiatry*, **33**, 281–300.
78. McLoughlin, D.M., Spargo, E., Wassif, W.S., *et al.* (1998). Structural and functional changes in skeletal muscle in anorexia nervosa. *Acta Neuropathologica*, **95**, 632–40.
79. Treasure, J.L., Fogelman, I., Russell, G.F.M., *et al.* (1987). Reversible bone loss in anorexia nervosa. *British Medical Journal*, **295**, 474–5.
80. World Health Organization. (1992). *International statistical classification of diseases and related health problems* (10th revision). WHO, Geneva.

81. Klibanski, A., Biller, B.M.K., Schoenfeld, D.A., *et al.* (1995). The effects of oestrogen administration on trabecular bone loss in young women with anorexia nervosa. *Journal of Clinical Endocrinology and Metabolism*, **80**, 898–904.
82. Serpell, L. and Treasure, J. (1997). Osteoporosis—a serious health risk in chronic anorexia nervosa. *European Eating Disorders Review*, **5**, 149–57.
83. Wakeling, A. and Russell, G.F.M. (1970). Disturbances in the regulation of body temperature in anorexia nervosa. *Psychological Medicine*, **1**, 30–9.
84. Rieger, W., Brady, J.P., and Weisberg, E. (1978). Hematologic changes in anorexia nervosa. *The American Journal of Psychiatry*, **135**, 984–5.
85. Russell, G.F.M. (1966). Acute dilation of the stomach in a patient with anorexia nervosa. *The British Journal of Psychiatry*, **112**, 203–7.
86. Beumont, P.J.V. and Large, M. (1991). Hypophosphataemia, delirium and cardiac arrhythmia in anorexia nervosa. *Medical Journal of Australia*, **155**, 519–22.
87. Russell, G.F.M., Treasure, J., and Eisler, I. (1998). Mothers with anorexia nervosa who underfeed their children: their recognition and management. *Psychological Medicine*, **28**, 93–108.
88. Russell, G.F.M. (1985). Premenarchal anorexia nervosa and its sequelae. *Journal of Psychiatric Research*, **19**, 363–9.
89. Lai, K.Y.C., de Bruyn, R., Lask, B., *et al.* (1994). Use of pelvic ultrasound to monitor ovarian and uterine maturity in childhood onset anorexia nervosa. *Archives of Disease in Childhood*, **71**, 228–31.
90. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
91. Beumont, P.J.V., Beardwood, C.J., and Russell, G.F.M. (1972). The occurrence of the syndrome of anorexia nervosa in male subjects. *Psychological Medicine*, **2**, 216–31.
92. Crisp, A.H. and Burns, T. (1990). Primary anorexia nervosa in the male and female: a comparison of clinical features and prognosis. In *Males with eating disorders* (ed. A.E. Andersen), pp. 77–99. Brunner–Mazel, New York.
93. Blake Woodside, D., Garfinkel, P.E., Lin, E., *et al.* (2001). Comparisons of men with full or partial eating disorders, and women with eating disorders in the community. *The American Journal of Psychiatry*, **158**, 570–4.
94. Burns, T. and Crisp, A.H. (1990). Outcome of anorexia nervosa in males. In *Males with eating disorders* (ed. A.E. Andersen), pp. 163–86. Brunner-Mazel, New York.
95. Fichter, M.M. and Daser, C. (1987). Symptomatology, psychosexual development and gender identity in 42 anorexic males. *Psychological Medicine*, **17**, 409–18.
96. Crisp, A.H., Norton, K., Gowers, S., *et al.* (1991). A controlled study of the effect of therapies aimed at adolescent and family psychopathology in anorexia nervosa. *The British Journal of Psychiatry*, **159**, 325–33.
97. Steinhausen, H.-C., Rauss-Mason, C., and Seidel, R. (1991). Follow-up studies of anorexia nervosa. A review of four decades of outcome research. *Psychological Medicine*, **21**, 447–54.
98. Morgan, H.G. and Hayward, A.E. (1988). Clinical assessment of anorexia nervosa. *The British Journal of Psychiatry*, **152**, 367–71.
99. Theander, S. (1985). Outcome and prognosis in anorexia nervosa and bulimia: some results of previous investigations, compared with a Swedish long-term study. *Journal of Psychosomatic Research*, **19**, 493–508.
100. Ratnasuriya, R.H., Eisler, I., Szmukler, G.I., *et al.* (1991). Anorexia nervosa: outcome and prognostic factors after 20 years. *The British Journal of Psychiatry*, **158**, 495–502.
101. Sullivan, P.F. (1995). Mortality in anorexia nervosa. *The American Journal of Psychiatry*, **152**, 1073–4.
102. Steinhausen, H.-C. and Seidel, R. (1993). Outcome in adolescent eating disorders. *The International Journal of Eating Disorders*, **14**, 487–96.
103. Wentz, E., Gillberg, C., Gillberg, I.C., *et al.* (2001). Ten-year follow-up of adolescent-onset anorexia nervosa: psychiatric disorders and overall functioning scales. *Journal of Child Psychology and Psychiatry*, **42**, 613–22.
104. Theander, S. (1992). Chronicity in anorexia nervosa: results from the Swedish long-term study. In *The course of eating disorders* (eds. W. Herzog, H.-C. Deter, and W. Vandereycken), pp. 198–213. Springer-Verlag, Berlin.
105. National Institute of Clinical Excellence. (2004). *Eating disorders: core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders*. Clinical Guideline, 9. National Collaborating Centre for Mental Health, London.
106. Halmi, K.A., Agras, S., Crow, S., *et al.* (2005). Predictors of treatment acceptance and completion in anorexia nervosa: implications for future study designs. *Archives of General Psychiatry*, **62**, 776–81.
107. Russell, G.F.M., Szmukler, G.I., Dare, C., *et al.* (1987). An evaluation of family therapy in anorexia nervosa and bulimia nervosa. *Archives of General Psychiatry*, **44**, 1047–56.
108. Lock, J., Agras, W.S., Bryson, S., *et al.* (2005). A comparison of short—and long-term family therapy for adolescent anorexia nervosa. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 632–9.
109. Eisler, I., Dare, C., Hodes, M., *et al.* (2000). Family therapy for adolescent anorexia nervosa: the results of a controlled comparison of two family interventions. *Journal of Child Psychology and Psychiatry*, **41**, 727–36.
110. Eisler, I., Dare, C., Russell, G.F.M., *et al.* (1997). Family and individual therapy in anorexia nervosa—a 5-year follow-up. *Archives of General Psychiatry*, **54**, 1025–30.
111. Russell, G.F.M. and Mezey, A.G. (1962). An analysis of weight gain in patients with anorexia nervosa treated with high calorie diets. *Clinical Science*, **23**, 449–61.
112. Eisler, I., Simic, M., Russell, G.F.M., *et al.* (2007). A randomized controlled treatment trial of two forms of family therapy in adolescent anorexia nervosa: a five-year follow-up. *Journal of Child Psychology and Psychiatry*.
113. Lock, J., Couturier, J., and Agras, W.S. (2006). Comparison of long-term outcomes in adolescents with anorexia nervosa treated with family therapy. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**, 666–72.
114. Fairburn, C.G., Shafran, R., and Cooper, Z. (1999). A cognitive-behavioural theory of anorexia nervosa. *Behaviour Research and Therapy*, **37**, 1–13.
115. Garner, D.M., Vitousek, K.M., and Pike, K.M. (1997). Cognitive-behavioural therapy for anorexia nervosa. In *Handbook of treatment for eating disorders* (2nd edn) (eds. D.M. Garner and P.E. Garfinkel), pp. 94–144. Guilford Press, New York.
116. Pike, K.M., Walsh, B.T., Vitousek, K., *et al.* (2003). Cognitive behaviour therapy in the posthospitalisation of anorexia nervosa. *The American Journal of Psychiatry*, **160**, 2046–9.
117. McIntosh, V.V.W., Jordan, J., Carter, F.A., *et al.* (2005). Three psychotherapies for anorexia nervosa: a randomized controlled trial. *The American Journal of Psychiatry*, **162**, 741–7.
118. Dare, C., Eisler, I., Russell, G., *et al.* (2001). Psychological therapies for adults with anorexia nervosa. *The British Journal of Psychiatry*, **178**, 216–21.
119. Freeman, C. (1992). Day patient treatment for anorexia nervosa. *British Review of Bulimia and Anorexia Nervosa*, **6**, 3–9.
120. Andersen, A.E., Bowers, W., and Evans, K. (1997). In-patient treatment of anorexia nervosa. In *Handbook of treatment for eating disorders* (2nd edn) (eds. D.M. Garner and P.E. Garfinkel), pp. 327–53. Guilford Press, New York.

121. Ramsay, R., Ward, A., Treasure, J., *et al.* (1999). Compulsory treatment in anorexia nervosa: short-term benefits and long-term mortality. *The British Journal of Psychiatry*, **175**, 147–53.
122. The Royal College of Psychiatrists. Eating Disorders. (1992). Council Report CR14, p. 13.
123. Touyz, S.O., Beumont, P.J.V., Glaun, D., *et al.* (1984). A comparison of lenient and strict operant conditioning programmes in refeeding patients with anorexia nervosa. *The British Journal of Psychiatry*, **144**, 517–20.
124. Prochaska, J.O. and Di Clemente, C.C. (1992). The transtheoretical approach. In *Handbook of psychotherapy integration* (eds. J.C. Norcross and M.R. Goldfried), pp. 300–34. Basic Books, New York.
125. Kaplan, A.S. and Olmsted, M.P. (1997). Partial hospitalisation. In *Handbook of treatment for eating disorders* (2nd edn) (eds. D.M. Garner and P.E. Garfinkel), pp. 354–60. Guilford Press, New York.
126. Crisp, A.H. (1997). Anorexia nervosa as flight from growth: assessment and treatment based on model. In *Handbook of treatment for eating disorders* (2nd edn) (eds. D.M. Garner and P.E. Garfinkel), pp. 248–77. Guilford Press, New York.
127. Mental Health Act Commission. (1997). *Guidance on the treatment of anorexia nervosa under the Mental Health Act 1983*, pp. 1–6. Mental Health Act Commission, Nottingham.
128. Stewart, D.A., Carter, J.C., Drinkwater, J., *et al.* (2001). Modification of eating attitudes and behaviour in adolescent girls: a controlled study. *The International Journal of Eating Disorders*, **29**, 107–18.

4.10.2 Bulimia nervosa

Christopher G. Fairburn,
Zafra Cooper, and Rebecca Murphy

Introduction

Origins of the concept

The history of the diagnosis bulimia nervosa begins as recently as 1979. It was in this year that Russell published his now seminal paper 'Bulimia nervosa: An ominous variant of anorexia nervosa'[1] in which he described 30 patients (28 women and 2 men), seen between 1972 and 1978, who had three major features in common. First, they had recurrent episodes of uncontrolled overeating; second, they regularly used self-induced vomiting or laxatives as means of weight control; and third, they had a morbid fear of becoming fat. Russell described many other features shared by these patients, including a history of anorexia nervosa (present in 80 per cent), the presence of severe depressive symptoms, and the fact that in most cases their body weight was in the healthy range. He noted that the disorder tended to run a chronic course and that it was 'extremely difficult to treat'. Finally, he suggested that this clinical picture should be viewed as a syndrome, distinct from anorexia nervosa, and he proposed the term 'bulimia nervosa'.

It is difficult to exaggerate the importance of Russell's paper. Its greatest contribution was perhaps its prescience in that it crystallized out from among the range of eating disorders seen in clinical practice a subgroup of patients that was just starting to become more common; it identified its central features; and it gave it an appropriate name.

Events since 1980

Events gathered pace in the 1980s. In 1980 a syndrome termed 'bulimia' was included in DSM-III.[2] This was intended to denote the type of patient that Russell had described, although its diagnostic criteria proved to be overly inclusive. In 1987, the criteria were refined and brought more in line with Russell's original concept. The syndrome was also renamed bulimia nervosa.[3] Also in the early 1980s evidence mounted that bulimia nervosa might be common and this led to a spate of studies of its prevalence. At the same time reports were published describing the successful treatment of these patients, the two most promising approaches being a specific form of cognitive behaviour therapy and the use of antidepressant drugs. By the mid-1980s, both treatments had been tested in the first of what has become a large series of controlled trials.

Now, three decades later, the diagnosis bulimia nervosa is included in both DSM-IV[4] and ICD-10,[5] its prevalence is established, aspects of its aetiology are beginning to be understood, and much has been learned about its treatment.

Classification and diagnosis

The classification of the eating disorders and their principal diagnostic criteria are shown in Table 4.10.2.1. Bulimia nervosa is one of the two main eating disorders recognized in DSM-IV and ICD-10, the other being anorexia nervosa (discussed in Chapter 4.10.1). In addition, in DSM-IV there is a residual category termed 'eating disorder not otherwise specified'. This is reserved for eating disorders of clinical severity that do not meet the diagnostic criteria for anorexia nervosa or bulimia nervosa.[6] In ICD-10, various eating disorder categories other than anorexia nervosa and bulimia nervosa are recognized (for example, atypical anorexia nervosa, atypical bulimia nervosa, overeating associated with other psychological disturbances), although these concepts have never been adequately defined or differentiated.

The relationship between the three DSM-IV diagnoses is represented diagrammatically in Fig. 4.10.2.1. The two overlapping inner circles represent anorexia nervosa (the smaller circle) and bulimia nervosa (the larger circle) respectively, the area of potential overlap being that occupied by those people who would meet the diagnostic criteria for both disorders but for the rule that the diagnosis of anorexia nervosa trumps that of bulimia nervosa (see Table 4.10.2.1). Surrounding these two circles is an outer circle which defines the boundary between having an eating disorder, a state of clinical significance, and having a lesser, non-clinical, eating problem. It is this boundary that demarcates what is, and is not, an eating disorder. Within the outer circle, but outside the two inner circles, lies eating disorder not otherwise specified (eating disorder NOS).

In DSM-IV a new eating disorder diagnosis was proposed termed 'binge eating disorder'. It is designed to denote an eating problem characterized by recurrent binge eating in the absence of the extreme weight-control behaviour seen in bulimia nervosa. Since binge eating disorder is a provisional diagnostic concept, it is currently an example of eating disorder NOS.

The two schemes for classifying eating disorders encourage the view that anorexia nervosa and bulimia nervosa are distinct clinical states. Consideration of their clinical features and course over time does not support this.[7] Binge eating disorder aside, patients with anorexia nervosa, bulimia nervosa, and eating disorder NOS have

Table 4.10.2.1 Classification of eating disorders and their principal diagnostic criteria.

Classification of eating disorders
◆ Anorexia nervosa
◆ Bulimia nervosa
◆ Eating disorder not otherwise specified (eating disorder NOS)
Binge eating disorder (a provisional new diagnosis, currently subsumed under eating disorder NOS)
Principal diagnostic criteria
Anorexia nervosa
1 Over-evaluation of shape and weight (i.e. judging self-worth largely, or exclusively, in terms of shape and weight)
2 Active maintenance of an unduly low body weight (e.g. body mass index < 17.5)
3 Amenorrhoea (in post-menarcheal females who are not taking an oral contraceptive). This criterion is often omitted.
Bulimia nervosa
1 Over-evaluation of shape and weight (i.e. judging self-worth largely, or exclusively, in terms of shape and weight)
2 Recurrent binge eating (i.e. recurrent episodes of uncontrolled overeating)
3 Extreme weight-control behaviour (e.g. strict dieting, frequent self- induced vomiting, or laxative misuse)
4 Diagnostic criteria for anorexia nervosa are not met
Eating disorder not otherwise specified
Eating disorders of clinical severity that do not conform to the diagnostic criteria for anorexia nervosa or bulimia nervosa
Binge eating disorder
Recurrent binge eating in the absence of the extreme weight-control behaviour seen in bulimia nervosa

(Reproduced from Fairburn, C.G. Cognitive Behaviour Therapy and Eating Disorders, copyright 2008, Guildford Press, NY.)

many features in common, most of which are not seen in other psychiatric disorders, and studies of their course indicate that patients migrate between these diagnoses over time: indeed, temporal migration is the norm rather than the exception. This temporal movement, together with the fact that the disorders share the same distinctive psychopathology, has led to the suggestion that the current diagnostic scheme is a poor reflection of clinical reality and that common 'transdiagnostic' mechanisms may be involved in the maintenance of eating disorder psychopathology.[8]

Clinical features

The great majority of patients with bulimia nervosa are female and most are in their 20s (although the age range is between 10 and 60 years). In considering the psychopathology of the disorder, a distinction may be drawn between its 'specific' and 'general' features. The former comprises features that are largely peculiar to eating disorders (for example, self-induced vomiting, the over-evaluation of shape and weight), whereas the latter consists of features seen in other psychiatric conditions (for example, depressive symptoms). The clinical features of bulimia nervosa are similar in men and women and in those with and without a history of anorexia nervosa.

Specific psychopathology

(a) Dieting and binge eating

The eating habits of patients with bulimia nervosa are characterized by strict dieting punctuated by repeated episodes of binge eating (see Fig. 4.10.2.2). The dieting is extreme and it is governed by multiple self-imposed dietary rules. These rules tend to be applied to all aspects of eating, including when to eat, what to eat, and how much to eat. As a result, the food eaten (when not binge eating) is restricted in quantity and range.

Recurrent episodes of 'binge eating' interrupt this dieting. (The term binge eating denotes discrete episodes of eating that have two characteristics: first, an unusually large amount of food is eaten,

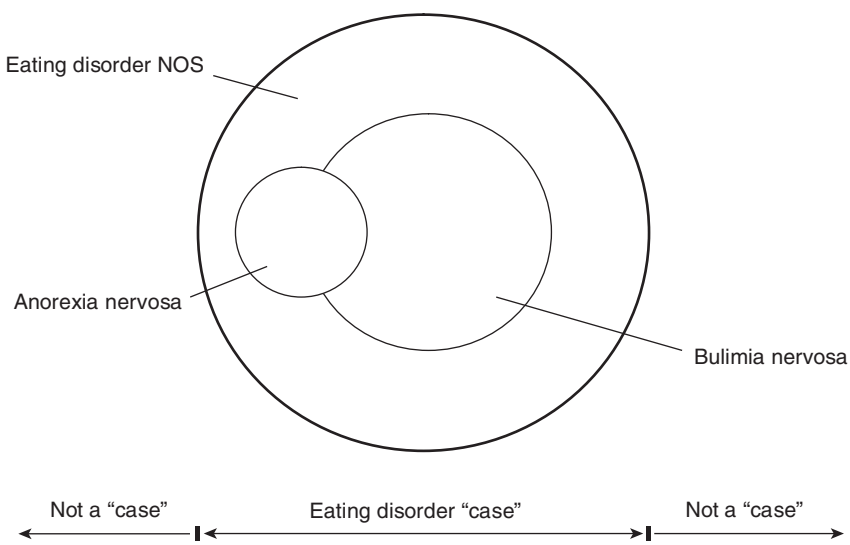


Fig. 4.10.2.1 A schematic representation of the relationship between anorexia nervosa, bulimia nervosa, and eating disorder NOS (Reprinted from Fairburn, C.G. and Bohn, K. Eating disorder NOS (EDNOS): an example of the troublesome "not otherwise specified" (NOS) category in DSM-IV, *Behaviour Research and Therapy*, **43**, 691–701, copyright (2005), with permission from Elsevier).

Day.....Thursday..... Date .....March 21 .....

Time	Food and drink consumed	Place	*	V/L	Context and comments
7.30	Glass water	Kitchen			[8stone 9lbs - really gross] Thirsty after yesterday
8:10	1 bowl muesli with skimmed milk Black coffee	Kitchen	*		Should have had less muesli. Must not binge today.
10:35	Half banana Black coffee	Work - at desk			Better - on track
11:45	Ham and lettuce slimline sandwich Diet coke	In canteen			Usual lunch
6.40 to 7.30	Slice of chocolate cake 1/2 tub ice cream 4 slices of toast with jam Diet coke Another slide of cake 2 slices of toast with jam Diet coke Jam from jar Two kit kats Mars bar Diet coke - large	Kitchen	*	V	Help - I can't stop eating. I'm completely out of control. I hate myself. I am disgusting. Why do I do this? I started as soon as I got in. I've ruined another day.
9:30	Rice cake with fat-free cheese Diet coke	Kitchen			Really lonely. Feel fat and unattractive. Feel like giving up.

Fig. 4.10.2.2 A monitoring record illustrating the eating habits of a patient with bulimia nervosa. Asterisks are used to signify episodes of eating that were viewed by the patient as excessive. The column headed 'V/L' is for recording episodes of self-induced vomiting or laxative misuse.

given the circumstances; and second, there is a sense of loss of control at the time. Some patients with eating disorders have similar episodes of uncontrolled overeating that do not involve the consumption of objectively large amounts of food. These episodes are sometimes referred to as 'subjective binges' although technically speaking they do not meet the definition of a 'binge'.) The frequency and regularity of the binge eating varies. Some patients have episodes almost every day, whereas in others the episodes are intermittent. In DSM-IV, it is specified that the binges should occur on average at least twice a week, but this is an arbitrary figure that has little discriminatory value. Among those patients in whom the binge eating is frequent, the binges have few, if any, obvious triggers, although there may be circumstances under which binge eating is more likely (for example, when alone at home). Among patients in whom the binge eating is less frequent, the binges often have clear precipitants. These tend to be of three overlapping types: first, there is breaking a personal dietary rule (for example, exceeding a daily calorie-limit or eating a banned food); second, there are situations which intensify concerns about shape and weight (for example, receiving an adverse comment about appearance); and third, there is the occurrence of negative moods (often as a result of interpersonal events). All three undermine the maintenance of strict dietary control.

The amount of food eaten during binges varies, both from patient to patient and from episode to episode. Typical episodes involve the consumption of 1000 to 2000 kcal.[9] The food eaten generally comprises items that are otherwise being avoided. Thus binges tend to be composed of energy-dense, high-fat items such as chocolate, ice cream, and pastries. Binges come to an end as a result of the combined influence of exhaustion, extreme fullness, a diminution of the drive to eat, and the running out of food supplies. In about three-quarters of patients they are immediately followed by measures designed to counteract the effects of the overeating, the most common being self-induced vomiting and the taking of laxatives or diuretics.

The binges are a source of considerable distress. They magnify these patients' fears of weight gain and fatness, and they may result in shame and self-disgust. For this reason most binges occur in private and are kept secret from others. It is the binge eating that eventually drives these people to seek help.

(b) Purging and other forms of weight control

In DSM-IV bulimia nervosa is subdivided into two types, a purging and non-purging type. In the purging type there is regular self-induced vomiting or the misuse of laxatives or diuretics, or both, whereas in the non-purging type such behaviour ('purging') is either not present or it is infrequent. The majority of patients seen in clinical practice have the purging form of the disorder and it has been the focus of most research.

Self-induced vomiting is the most common form of purging. In most patients it only takes place after binge eating. It is generally achieved by stimulating the gag reflex, using the fingers or some

other long object, although in more established cases it can be accomplished with no mechanical aid. The vomiting is repeated until patients think that they have retrieved all the food that they can. Patients get extremely distressed if they are unable to vomit after binge eating: indeed, if they foresee that they may not have the opportunity to vomit, they tend not to binge. A minority of patients also induce vomiting at other times, for example, following smaller episodes of overeating (subjective binges) or ordinary meals or snacks.

The misuse of laxatives or diuretics is somewhat less common than self-induced vomiting. It takes two forms: one is to compensate for specific episodes of binge eating, like self-induced vomiting; and the other is as a general method of weight control (like dieting), in which case it is not tied to particular episodes of overeating. The number of laxatives or diuretics taken varies considerably, sometimes far exceeding the recommended dose.

None of these methods of purging is an effective method of weight control. Self-induced vomiting results in the retrieval of only about half to two-thirds of what has been eaten, the taking of laxatives has a minimal effect on food absorption, and diuretic-taking has none. As a result, a significant proportion of each binge is absorbed.

The weight of most of these patients is in the healthy range (BMI between 20 and 25) due to the effects of the under-eating and over-eating cancelling each other out. As a result they do not experience the secondary psychosocial and physical effects associated with maintaining a very low weight seen in anorexia nervosa.

Other forms of weight-control behaviour are practised by some patients, including over-exercising, the spitting out of food, and the taking of repeated enemas or saunas. Over-exercising is the most common of these, but it is not nearly as prominent or as obviously pathological as in anorexia nervosa. A minority of patients ruminate, that is, repeatedly regurgitate and re-chew food that has been eaten. They may then either re-swallow the food or spit it out. This behaviour is not well-understood.

In the non-purging type of bulimia nervosa there is no vomiting or misuse of laxatives or diuretics, or they occur infrequently. Instead, there is sustained and marked dietary restriction outside the binges. This is both a response to the binge eating and contributor to it, in that this type of eating increases the risk of further episodes. In all other respects the two subtypes of the disorder are similar.

(c) Attitudes to shape and weight

A characteristic set of attitudes to shape and weight is the other distinctive element of the specific psychopathology of bulimia nervosa. Equivalent attitudes are found in anorexia nervosa and most cases of eating disorder NOS. These attitudes are often described as the 'core psychopathology' of eating disorders. They are characterized by an overconcern with shape and weight in which there is a fear of weight gain and fatness that is generally accompanied by a pursuit of weight loss and thinness. Underlying this psychopathology is the tendency to judge self-worth largely, or even exclusively, in terms of shape and weight. Whereas it is usual to evaluate self-worth on the basis of perceived performance in a variety of domains of life (such as interpersonal relationships, work, sport, artistic ability, etc.), people with anorexia nervosa or bulimia nervosa evaluate themselves primarily in terms of their shape and weight. These attitudes and values constitute a good example of an overvalued idea.

Most features of bulimia nervosa can be understood as being secondary to these attitudes to shape and weight. The dieting, purging, and over-exercising are obvious secondary features. In addition, there are direct behavioural expressions of these concerns. For example, many patients repeatedly weigh themselves and scrutinize their appearance in mirrors. Others avoid any knowledge of their weight while being acutely sensitive about their appearance. Some avoid others seeing their body and some even avoid seeing it themselves. This can have a major impact on social and sexual relationships.

The concerns about shape and weight, and eating, have a major effect on others in the patient's immediate environment. Meals are often times of tension and social events which involve eating may be avoided. The feeding of children may be affected[10] and their growth may be impaired[11] (see Chapter 9.3.6).

General psychopathology

General psychiatric symptoms are prominent in bulimia nervosa; more so than in anorexia nervosa. The nature of the comorbid symptoms also differs. Depressive features are particularly striking: indeed, the level of depressive symptoms in bulimia nervosa is equivalent to that seen in major depressive disorder. Anxiety symptoms are also encountered, many of which are directly related to the eating disorder; for example, there is often pronounced anxiety about eating in public. Obsessive-compulsive features are sometimes present, although they are less common than in anorexia nervosa. Similarly, social functioning is less impaired.

The depressive features of bulimia nervosa deserve special mention. In most patients the depressive features can be attributed to the presence of the eating disorder but in a subgroup there appears to be an independent coexisting, but interacting, clinical depression. Features suggestive of such coexisting clinical depressions include the following: recent intensification of depressive features (in the absence of any change in the eating disorder or the patient's circumstances); pervasive and extreme negative thinking (i.e. broader in content than concerns about eating, shape, and weight); hopelessness in general (i.e. seeing the future as totally bleak, seeing no future, resignation); recurrent thoughts about death and dying; suicidal thoughts; guilt over events in the far past; a decrease in involvement with others over and above any impairment that already accompanied the eating disorder (e.g. ceasing to see friends); loss of interest in activities that had been pursued despite the eating disorder (e.g. ceasing to listen to music; ceasing to read newspapers or follow the news); and a decrease in drive and initiative.

These coexisting clinical depressions often go undetected since they are viewed as characteristic of bulimia nervosa. This is unfortunate for two reasons: first, they interfere with the treatment of the eating disorder; and second, they are readily treated with antidepressant drugs (unlike the secondary depressive features).

A minority of patients with bulimia nervosa have 'impulse-control' problems, such as the overconsumption of alcohol or drugs, or repeated self-harm (e.g. cutting). Some of these patients also meet diagnostic criteria for borderline personality disorder (see Chapter 4.12.2). The prevalence of such features varies according to treatment setting: they are unusual in community samples, whereas they are more frequent among patients seen in specialist treatment centres.

Much more common than frank personality disorders are two traits which are also seen in anorexia nervosa. The first is low self-esteem. This generally antedates the eating disorder, although it is often exaggerated by it. Many patients with bulimia nervosa describe longstanding doubts about their worth and ability, irrespective of their accomplishments. The second is perfectionism, that is, imposing on oneself inordinately high personal standards in a range of domains (for example, work, sport, personal conduct). Since many of these standards are unachievable, it is common for these patients to give long histories of viewing themselves as perpetually failing.

Physical features

There are few physical abnormalities in bulimia nervosa. Body weight is unremarkable in the majority of patients and thus the physical effects of starvation are rarely seen. Nevertheless, menstrual abnormalities or amenorrhoea are present in about a quarter of patients. These are likely to be secondary to the disturbed eating since they generally respond to the successful correction of the eating disorder. On laboratory testing endocrine abnormalities are sometimes encountered and these tend to be mild versions of those found in anorexia nervosa. Fertility appears not to be affected.

Frequent purging, and especially the combination of vomiting and laxative misuse, results in fluid and electrolyte abnormalities in some patients. These abnormalities are varied in nature but most often consist of some combination of hypokalemia, hyponatremia, and hypochloremia. The patients appear to accommodate to these abnormalities since medically serious complications (for example, cardiac arrhythmias) are much less common than might otherwise be expected given the laboratory findings. Some patients experience intermittent oedema particularly if there is a sudden decrease in the extent of their purging.

Localized physical abnormalities include erosion of the dental enamel (especially from the lingual surfaces of the front teeth) among those who have vomited for many years; traumatic calluses on the knuckles of the hand of those who use their fingers to induce the gag reflex (Russell's sign); and enlargement of the salivary glands, especially the parotids, probably as a result of chronic inflammation. A small proportion of patients have raised serum amylase levels usually due to an increase in the salivary isoenzyme.

Relationship to other disorders

Anorexia nervosa and eating disorder NOS

Bulimia nervosa has many features in common with anorexia nervosa and eating disorder NOS, particularly the characteristic attitudes to shape and weight and the behaviour that arises directly as a result.[12] In most cases, bulimia nervosa is preceded either by frank anorexia nervosa (in about a quarter of cases) or an anorexia nervosa-like form of eating disorder NOS. While movement from bulimia nervosa to anorexia nervosa is unusual, progression on to some form of eating disorder NOS is common. Whether it is appropriate to view such patients as having recovered from one psychiatric disorder and developed another is a moot point: rather, it would seem more appropriate to view them as having a single evolving eating disorder.

There is some evidence of co-aggregation between bulimia nervosa, anorexia nervosa, and eating disorder NOS with there being increased rates of all three diagnoses among the relatives of probands with either condition.[13]

Obesity

Few patients with bulimia nervosa are overweight or have obesity. On the other hand there is evidence of raised rates of parental and premorbid obesity.[14] Obesity is an unusual sequel of the disorder although this may be because those at most risk of obesity are less likely to recover and so continue to suppress their weight.

Other psychiatric disorders

As noted above, depressive features are common in bulimia nervosa and they may antedate the eating disorder. The same is true of anxiety and anxiety disorders. Most family studies have found a raised rate of affective disorder among these patients' relatives whereas little is known about the familial relationship between bulimia nervosa and the anxiety disorders.[15] There is a raised rate of alcohol and drug abuse among patients with bulimia nervosa and a raised rate among these patients' relatives.[15] Substance abuse rarely antedates the eating disorder but this is to be expected given the age of onset of substance abuse disorders.

It is hazardous making personality disorder diagnoses among those with eating disorders. This is because eating disorders have their onset in adolescence and they directly affect many of the characteristics upon which personality is judged. Thus there is a risk of overestimating the presence of personality disturbance. Nevertheless, some patients with bulimia nervosa do seem to have a coexisting personality disorder, the most common form being borderline personality disorder. Little is known about the rate of personality disturbance among these patients' relatives although there is evidence of familial co-aggregation of anorexia nervosa and obsessional and perfectionist traits.[15]

Diabetes mellitus

It was thought that eating disorders were over-represented among those with Type I diabetes mellitus. This is now not clear. Controlled studies in which eating disorder features have been assessed by interview rather than self-report questionnaire (the preferred method and one which minimizes the risk of false positive diagnoses) have found little evidence of an elevated rate of anorexia nervosa although the rate of bulimia nervosa may be increased.[16]

Distribution

The fact that it took Russell more than 6 years (1972–1978) to collect 30 cases of bulimia nervosa suggested that the disorder was not common. It is therefore remarkable that within a few years of the publication of Russell's paper it was evident that bulimia nervosa was an important source of psychiatric morbidity among young women.

In the early 1980s large numbers of previously undetected cases were identified using the media.[17,18] These cases were remarkably similar to Russell's, except that almost all were female and a small proportion had a history of anorexia nervosa. Most had kept their eating disorder secret for many years, and because of shame and hopelessness few had sought help. Many thought that they were the only person with this type of eating disorder. Simultaneously however, and doubtless partly as a result of the media attention,

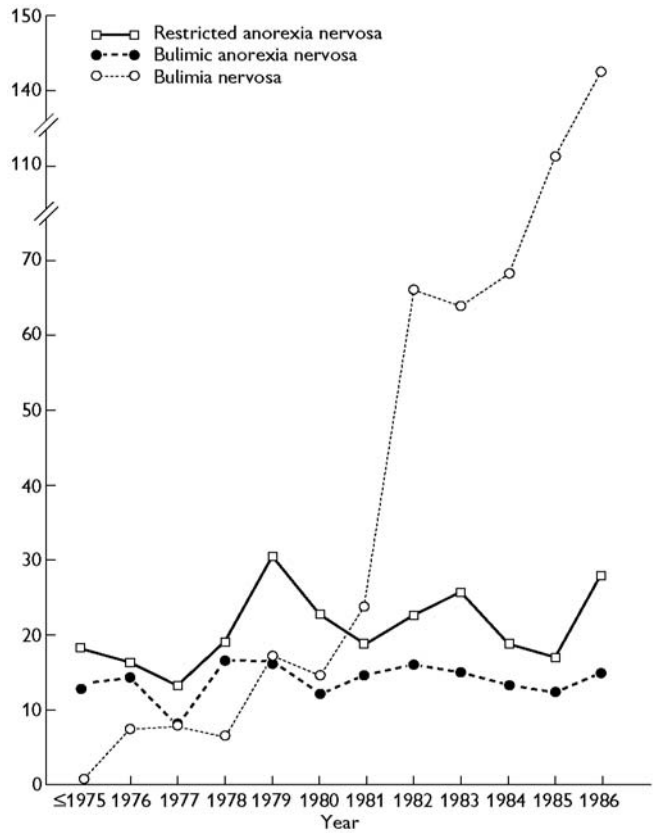


Fig. 4.10.2.3 Rates of referral to a major eating disorder centre in Toronto (1975–1986) (Reproduced from Garner, D.M. and Fairburn, C.G Relationship between anorexia nervosa and bulimia nervosa: diagnostic implications. In *Diagnostic issues in anorexia nervosa and bulimia nervosa* (eds. D.M. Garner and P.E. Garfinkel), copyright 1998, Brunner/Mazel, New York).

there was also a sharp increase in the number of people requesting treatment for bulimia nervosa (see Fig. 4.10.2.3).

The marked increase in the number of patients with bulimia nervosa stimulated interest in the prevalence of the disorder. By 1989 over 50 prevalence studies had been conducted, many of which yielded unrealistically high prevalence figures as a result of using weak assessment and sampling procedures. Gradually methods improved with the result that estimates of the prevalence of bulimia nervosa decreased to more modest and consistent levels with the point prevalence among young adult women (aged 16 to 35 years) being in the region of 1 per cent,[19,20] a similar figure being obtained for lifetime prevalence.[20] The prevalence of bulimia nervosa among men is not known. Among patient samples, male cases are unusual. Bulimia nervosa is thought to be uncommon in non-Western societies although few prevalence studies have been conducted and most have had significant methodological shortcomings.[21]

There have been few estimates of the incidence of bulimia nervosa, and since these have been based upon clinic rather than community samples, they are likely to underestimate the true figures. Even today, many people with bulimia nervosa do not seek help. The lack of reliable community-based incidence figures also makes it impossible to know whether the disorder has become more common since the 1970s or whether the upsurge in cases in the early 1980s was as a result of undetected cases being more likely to seek treatment. Data from the assessment of women in different birth cohorts suggest that the disorder has become much more common[22] although other explanations for the apparent increase cannot be ruled out.

Aetiology

Development of the disorder

As noted in Chapter 4.10.1, anorexia nervosa generally starts in mid-adolescence with a period of voluntary dietary restriction which proceeds to get out of control. As a result body weight falls and a state of starvation develops. Shape and weight concerns may pre-date the onset of the dieting or develop as weight is lost.

Bulimia nervosa starts in a similar way although the age of onset is typically some years later and shape and weight concerns usually antedate the dieting. The dietary restriction resembles that seen in anorexia nervosa and it leads to weight loss sufficient to result in anorexia nervosa in about a quarter of cases. (As a result of referral bias, this proportion is higher in cases seen in specialist centres.) In the remaining cases there is also weight loss but it is less extreme. After a variable length of time (generally within 3 years) dietary control breaks down with the patient's dieting becoming punctuated by episodes of overeating. At first, the episodes of overeating may be modest in size and intermittent, but gradually they become larger and more frequent. As a result, the lost weight is regained and body weight returns to near its original level. By this point the disorder tends to be self-perpetuating. At some stage in this sequence of events, self-induced vomiting and laxative misuse may be adopted to compensate for the overeating. In practice, however, both forms of behaviour have the opposite effect since belief in their effectiveness encourages a relaxation of control over eating. In those who vomit this phenomenon is exaggerated by the discovery that the process is easier after eating large amounts of food.

Predisposing factors and processes

There are many risk factors for the development of bulimia nervosa[14] and these overlap with those for anorexia nervosa[23] (see Table 4.10.2.2).

The risk factors may be usefully divided into a number of categories.

- Demographic factors—these are being female, adolescent and living in a Western society.
- Exposure to an immediate social environment that encourages dieting—this includes being brought up in a family in which there is intense interest in shape, weight, and eating as a result of one or more members either having some degree of eating disorder or having a medical condition that affects eating or weight (such as diabetes mellitus). Extreme occupational or recreational pressures to diet also appear to be associated with increased risk (for example, ballet dancing), although there may also be an element of self-selection. Another important influence is parental and childhood obesity, the rates of which are substa ntially increased in bulimia nervosa. Both are likely to sensitize individuals to their appearance and weight, and thereby make them prone to diet. There is also some evidence that puberty occurs comparatively early which may also magnify concerns about shape.

Table 4.10.2.2 Principal risk factors for anorexia nervosa and bulimia nervosa

General factors
Female
Adolescence and early adulthood
Living in a Western society
Individual-specific factors
Family history
Eating disorder of any type
Depression
Substance abuse, especially alcoholism (bulimia nervosa)
Obesity (bulimia nervosa)
Premorbid experiences
Obstetric complications
Adverse parenting (especially low contact, high expectations, parental discord)
Sexual abuse
Family dieting
Critical comments about eating, shape, or weight from family and others
Occupational and recreational pressure to be slim
Premorbid characteristics
Low self-esteem
Perfectionism (anorexia nervosa and to a lesser extent bulimia nervosa)
Neuroticism
Anxiety and anxiety disorders
Obesity (bulimia nervosa)
Early menarche (bulimia nervosa)
Type I diabetes (bulimia nervosa)

- Exposure to factors that increase the risk of psychiatric disturbance in general and depression in particular—these include a family history of psychiatric disorder, especially depression, and a range of adverse childhood experiences including parenting deficits and sexual and physical abuse. It was thought that sexual abuse was especially common among those who develop bulimia nervosa, but the balance of evidence suggests that the rate is no higher than that among those who develop other psychiatric disorders.
- Perfectionism and low self-esteem—both traits are common antecedents of anorexia nervosa and bulimia nervosa. Typically they interact resulting in feelings of incompetence and ineffectiveness.
- Family history of substance abuse—there is a raised rate of substance abuse in the families of patients with bulimia nervosa. It is not clear how this increases the risk of bulimia nervosa. Clinical observations suggest that some of those who develop bulimia nervosa learn to modulate their mood by engaging in self-harm (e.g. by cutting themselves) or by consuming large quantities of food, alcohol, or psychoactive drugs.

An important question is how those with anorexia nervosa are able to maintain strict control over their eating whereas this is not true of those with bulimia nervosa. The explanation is unclear but several processes may be of relevance. First, perfectionism is even more pronounced in anorexia nervosa and it may enhance self-control. Second, the vulnerability to obesity found in bulimia nervosa may somehow undermine dietary restraint. Third, the mood lability of bulimia nervosa may also disrupt restraint.

Contribution of genetic factors

The fact that eating disorders run in families suggests that there may be a genetic contribution. In the absence of adoption studies, twin designs have been used to establish its extent.[24] Clinic samples show concordance for anorexia nervosa of around 55 per cent in monozygotic twins and 5 per cent in dizygotic twins, with the corresponding figures for bulimia nervosa being 35 and 30 per cent, respectively.[7,25] These findings suggest a significant heritability to anorexia nervosa but not to bulimia nervosa. Despite this, there is uncertainty as to the size of the genetic contribution to both disorders with there being differing point estimates and wide confidence intervals. The same applies to the contributions of individual-specific and shared (common) environmental factors. A number of issues affect the interpretation of the data. For example, there has been insufficient power to detect shared environmental effects, and established diagnostic criteria have been broadened considerably to increase the number of 'affected' twins available for analysis.

Given the clear and possibly substantial genetic contribution to both disorders, molecular genetic studies have been conducted to identify the underlying loci and genes. Genetic association studies have focussed in particular on polymorphisms in 5-HT (serotonin)-related genes because this neurotransmitter system is important in regulation of eating and mood, but a range of other polymorphisms have also been investigated. Despite this, no associations with eating disorders have been clearly replicated or confirmed in a family study or by meta-analysis. There has been one multi-centre genome-wide linkage study. It found linkage peaks for anorexia nervosa and bulimia nervosa on chromosomes 1, 4, 10, and 14. A further analysis, which covaried for related behavioural traits, identified a different locus on chromosome 1, as well as loci on chromosomes 2 and 13. All these findings await replication.

Neurobiological findings

There has been extensive research into the neurobiology of eating disorders. This has focussed on neuropeptide and monoamine (especially 5-HT) systems thought to be central to the physiology of eating and weight regulation.[26] Of the various central and peripheral abnormalities reported, many are likely to be secondary to the disturbed eating and associated weight loss. However, some aspects of 5-HT function and its receptors remain abnormal after recovery, leading to speculation that there is a trait monoamine abnormality which may predispose to the development of eating disorders or to associated characteristics such as perfectionism. Furthermore, normal dieting in healthy women has been shown to alter central 5-HT function, providing a potential mechanism by which eating disorders might be precipitated in women vulnerable for other reasons.

Brain functional imaging studies have identified altered activity in the frontal, cingulate, temporal, and parietal cortical regions in both anorexia nervosa and bulimia nervosa, and there is some evidence that these alterations persist after recovery.[27] Whether they are a consequence of the eating disorder or have somehow contributed to it is not known.

Maintaining factors and processes

Once established, bulimia nervosa tends to run a chronic course although it commonly evolves into eating disorder NOS. There are a number of processes which account for its self-perpetuating character. These are discussed in Chapter 6.3.2.2 (CBT for eating disorders). They include the ongoing influence of the extreme concerns about shape and weight; the form of these patients' dieting, which encourages binge eating; the mood-modulating effect of binge eating; and the fact that the loss of control overeating perpetuates fears of weight gain.

Assessment

The identification of patients with bulimia nervosa is not difficult so long as the diagnosis is considered. This is important because the shame that characterizes the disorder leads many people to delay seeking help—the average delay between onset and presentation is about 5 years—and when they do present for treatment, some do so indirectly. Thus they may complain of depression, substance abuse, menstrual disturbance, or gastrointestinal symptoms, rather than the eating disorder itself. The best policy is for psychiatrists to always keep in mind the possibility of bulimia nervosa when assessing female patients aged between 16 and 35 years. Negative responses to the following questions should suffice to exclude the disorder:

- 'Do you have any problems with your eating?'
- 'Do you have any problems controlling your eating; that is, problems with binge eating?'
- 'Do you ever make yourself sick or take laxatives to control your weight or shape?'

Patients who present directly complain of having lost control over eating and their assessment is generally straightforward. It should always include an assessment of the extent to which the disorder is interfering with everyday functioning and an evaluation of general psychiatric features and especially those of depression.

The best established measure of eating disorder features is the Eating Disorder Examination.[28] This interview is widely regarded as the 'gold standard' measure of eating disorders, but it is possibly too exhaustive to use on a routine clinical basis. Various self-report questionnaires are available but they provide a more basic level of assessment and they cannot be used to make a clinical diagnosis. The leading self-report measures are the Eating Disorder Inventory[29] and the self-report version of the Eating Disorder Examination.[30]

No assessment is complete without weighing the patient and checking their height. Weighing needs to be done with considerable sensitivity because of these patients' concerns about their weight. A physical examination is not essential unless the patient is underweight (or there are other medical indications), nor are laboratory tests required except in those cases in which there is reason to suspect that there might be fluid or electrolyte disturbance.

Treatment

Given that bulimia nervosa has only recently been described, there has been a remarkable amount of research on its treatment. Over 70 randomized controlled trials have been completed. An authoritative meta-analysis has been conducted by the United Kingdom National Institute for Health and Clinical Excellence or 'NICE'.[31] The majority of the trials have focussed on adults with bulimia nervosa, the treatment of adolescents having received little attention, and almost all these studies have been 'efficacy' rather than 'effectiveness' studies. However, there are reasons to think that their findings are of direct relevance to routine patient care not least because the patients studied have been similar to those seen in clinical practice. Nevertheless, there is a definite need for effectiveness studies particularly now that the main treatment options are clear.

Studies of pharmacological treatment

A variety of drugs have been tested as possible treatments for bulimia nervosa including antidepressants, appetite suppressants, anticonvulsants, and lithium. Only antidepressants have shown promise.

(a) Antidepressant medication

All the major classes of antidepressant drug have been evaluated, including tricyclic antidepressants, monoamine oxidase inhibitors, selective serotonin uptake inhibitors, and atypical antidepressants. The findings have been relatively consistent and may be summarized as follows (adapted from Wilson and Fairburn[32]):

- Antidepressant drugs are more effective than placebo at reducing the frequency of binge eating and purging. On average, among treatment completers there is about a 50 per cent reduction in the frequency of binge eating and a cessation rate of about 20 per cent. The therapeutic effect is more rapid than that seen in depression. There is generally little change in the placebo group. The dropout rate varies but averages about 30 per cent.
- The longer-term effects of antidepressant drugs remain largely untested. Almost all the studies to date have been of their short-term use (16 weeks or less). The findings of the few longer-term studies suggest that outcome is poor and compliance low.[33]
- Few studies have evaluated the effects of antidepressant drugs on features other than binge eating and purging. Mood improves as the frequency of binge eating declines but this effect is common to all treatments for bulimia nervosa. Antidepressant drugs do not appear to modify these patients' extreme dieting which may account for the poor maintenance of change.
- Different antidepressant drugs seem to be equally effective, although there have been no direct comparisons of different drugs.
- With one exception, there have been no systematic dose-response studies. The exception showed that fluoxetine at a dose of 60 mg/day, but not 20 mg/day, was more effective than placebo.[34]
- Patients who fail to respond to one antidepressant drug may respond to another. There have been no drug augmentation studies.
- No consistent predictors of response have been identified. Pretreatment levels of depression appear not to be related to outcome.

- The mechanism(s) whereby antidepressant drugs exert their 'antibulimic' effects is not known. The apparent comparability of different classes of drug implicates a common mechanism but this is unlikely to be their antidepressant action since the response is too rapid and the level of depression does not predict outcome.

Studies of psychological treatment

(a) Cognitive behaviour therapy (CBT)

The most intensively studied psychological treatment is a specific form of CBT.[35,36] This was the first promising treatment described and it remains the leading treatment for the disorder. The treatment and its rationale are described in Chapter 6.3.2.2.

CBT is conducted on an outpatient basis and involves 15 to 20 sessions over about 5 months. It is suitable for all patients bar the small minority (less than 5 per cent) who require hospitalization.

The findings of the studies of CBT (over 20 controlled trials) are summarized below (adapted from Wilson and Fairburn[32]):

- The drop-out rate with CBT (about 20 per cent) is less than that seen with antidepressant drugs. The treatment is also more acceptable to these patients than treatment with medication.
- CBT has a substantial effect on the frequency of binge eating and purging. On average, among treatment completers there is about an 80 per cent reduction in the frequency of binge eating, and a cessation rate of about 60 per cent.
- The effects of CBT appear to be well-maintained. Most of the recent studies have included a 6 to 12-month follow-up period. The relapse rates are low.
- CBT affects most aspects of the psychopathology of bulimia nervosa including the binge eating, purging, dietary restraint, and the over-evaluation of shape and weight. In common with other treatments, the level of depression decreases as the frequency of binge eating declines. Social functioning and self-esteem also improve.
- CBT is more effective than delayed treatment (i.e. a waiting list control group), other psychological treatments (other than possibly interpersonal psychotherapy—see below) and antidepressant drugs at reducing the frequency of binge eating and purging. Among the other psychological treatments studied have been supportive psychotherapy, focal psychotherapy, supportive-expressive psychotherapy, interpersonal psychotherapy, hypnobehavioural treatment, stress management, nutritional counselling, behavioural versions of cognitive behaviour therapy, and exposure with response prevention.
- No consistent predictors of response to CBT have been identified. Severity of symptoms at presentation, a history of anorexia nervosa, low self-esteem, and the presence of borderline personality disorder have been associated with worse outcome in some studies but not others. However, the extent of initial response (over the first 4 weeks of treatment) is a potent and potentially valuable predictor of outcome.[37,38]
- The mechanism(s) of action of CBT have yet to be established although they appear to be mediated at least in part by a reduction in dietary restraint.[39] It also seems that the cognitive procedures are required for progress to be maintained since behavioural versions of the treatment are associated with a greater risk of relapse.[40]
- There is some evidence that the combination of CBT and antidepressant drugs may be more effective than CBT alone in reducing accompanying anxiety and depressive symptoms.
- Simpler forms of CBT may help a small subset of patients although the findings are not consistent. These include brief versions of the treatment and cognitive behavioural self-help in which patients follow a cognitive behavioural self-help programme either on their own (pure self-help) or with the guidance of a therapist (guided self-help). The programme may be delivered via a book, CD-ROM or the internet. This type of treatment is still in its infancy and it has yet to be rigorously evaluated.

(b) Interpersonal psychotherapy (IPT)

IPT is the leading alternative to CBT. This treatment was originally devised by Klerman and colleagues as a treatment for depression (see Chapter 6.3.3).[41] It is a focal psychotherapy, the main emphasis of which is to help patients identify and modify current interpersonal problems. The treatment is both non-directive and non-interpretative and, as adapted for bulimia nervosa,[42] it pays little attention to the patient's eating disorder. It is therefore very different to CBT. There have been two comparisons of CBT and IPT in the treatment of bulimia nervosa and both have found that they are about comparably effective but that IPT takes 4 to 8 months longer to achieve its effects.[40,43] The second of these studies was also designed to identify variables that might allow patients to be matched to CBT or IPT but none emerged.

Management of bulimia nervosa

Given that the leading treatment for bulimia nervosa is CBT, the ideal form of management is the provision of CBT by a therapist trained in its implementation. Unfortunately there is a shortage of the necessary expertise with the consequence that a 'stepped care' approach has been advocated on pragmatic grounds. With such an approach a simple treatment is used first and only if this proves insufficient a more complex and specialized intervention is provided. Three steps may be distinguished.

Step 1—Having established the diagnosis, the first decision is whether the patient may be managed on an outpatient basis. The great majority (over 95 per cent of referrals to non-specialist centres) may be managed this way. Exceptions are patients at significant risk of suicide and the presence of physical complications necessitating inpatient or day patient care. Severe substance abuse requires treatment in its own right, although this can sometimes be integrated with the treatment of the eating disorder. For example, it is possible to adapt CBT for bulimia nervosa so that it addresses the patient's substance abuse at the same time.

Step 2—If the patient is suitable for outpatient-based treatment, guided cognitive behavioural self-help, and/or antidepressant medication is the next step. The former involves following a cognitive behavioural self-help programme under the guidance of a 'facilitator' (a non-specialist therapist). Three cognitive behavioural self-help books are available,[44–46] two of which are direct translations of CBT for bulimia nervosa. There is evidence to support the use of all three and it is a matter of preference which is chosen. Each provides information about bulimia nervosa together with a self-help programme. The role of the facilitator is not to provide

treatment as such, as in a conventional 'therapist-led' treatment, but rather to support and encourage the patient to follow the programme. Thus, this is a 'programme-led' form of treatment, and it is this characteristic that makes it suitable for widespread dissemination. Treatment generally takes about 4 months and involves eight to ten meetings with the facilitator, each lasting up to 30 min. It is best if the first few appointments are weekly.

Guided self-help may take place in a variety of settings including primary care. Unfortunately only a small proportion of patients show substantial change and there are no reliable predictors of outcome. Patients who have obtained little benefit after 4 to 6-weeks are unlikely to do so and should be moved on to Step 3.

Antidepressant medication is an alternative to guided self-help. It is important to note that the medication is being used for its antibulimic effect not its antidepressant action. This affects the choice of drug and the dose chosen. The drug of choice is fluoxetine at a dose of 60 mg in the morning. It is usually well tolerated. As with guided self-help, if there has been little benefit after 4 to 6-weeks it is best to move on to Step 3.

A third alternative would be to combine antidepressant medication and guided self-help to see if the two augment each other. This would be a reasonable strategy although there are few data to support it.

Step 3—Patients who do not benefit from cognitive behavioural self-help or antidepressant medication should receive full CBT on a one-to-one basis (see Chapter 6.3.2.2). Ideally this should be delivered by a well-trained therapist who is used to following the protocol.[36]

Step 4—The fourth step is for those who are still symptomatic after having received well-delivered CBT. This step is pragmatic since there are few research findings of relevance. To guide the choice of treatment, the reasons for the poor response need to be carefully considered. Explanations include the presence of an undetected clinical depression; failure of CBT (which itself needs to be explained); poorly delivered CBT; poor patient compliance (which also needs to be explained); and disruption by outside events.

There are a number of different treatment options under these circumstances, the choice depending upon the outcome of the reassessment and the resources available. They include the following:

- Stop treatment and arrange to re-evaluate the patient after an interval of some months. Some patients and therapists 'burn out' after a sustained period of therapeutic work. A break can often be helpful. Deciding to stop treatment should be a joint decision and it is not appropriate with patients who are distressed or with those whose physical or psychological well-being is a cause for concern.
- Embark upon a new psychological treatment. While the obvious choice is IPT, there are no grounds for supposing that patients who fail to respond to CBT will respond to IPT. Indeed, there is evidence that this is not the case.[47] An alternative strategy would be to change the form of the CBT. The re-evaluation of the patient may have resulted in the identification of problems that might be amenable to cognitive behavioural procedures outside the realm of mainstream CBT for bulimia nervosa. Indeed, the fact that this occurs has led to the development of a new 'enhanced' form of CBT for eating disorders that is not only designed to be more potent that the earlier treatment but it also addresses common obstacles to change external to the eating disorder psychopathology (i.e. mood intolerance, clinical perfectionism, low self-esteem, and interpersonal problems).[8,48] It is also of note that this treatment is designed to be suitable for all forms of eating disorder not just bulimia nervosa.
- Arrange for day patient or inpatient treatment. In a small minority of cases outpatient treatment proves not to be sufficient, either because the disorder is resistant to outpatient-based forms of treatment or because the patient's life circumstances are interfering with progress. In such cases day patient or inpatient treatment can be useful. Generally this involves a combination of therapeutic approaches including elements of CBT. It is essential that both day patient treatment and inpatient treatment are followed by outpatient treatment designed to ensure that progress is maintained.

Course and outcome

Much remains to be learned about the course and outcome of bulimia nervosa. It is clear from epidemiological studies that many people do not present for treatment. The course of their disorder is completely unknown. Those who do present tend to do so after a considerable period of time indicating that among this subgroup the disorder has a tendency to run a protracted course. On the other hand, the findings of the treatment studies indicate that the outcome is considerably better than Russell originally suggested, although it must be stressed that even with CBT, the most effective treatment, only about half the patients make a full and lasting recovery.

There have been few studies of long-term course or outcome. A 5-year prospective study of a community sample found that at each assessment point between a half and two-thirds of the cases had an eating disorder of clinical severity, the majority being cases of eating disorder NOS.[49] A 10-year follow-up study found that about 10 per cent met diagnostic criteria for bulimia nervosa and a further 20 per cent had a form of eating disorder NOS.[50] There is no evidence that bulimia nervosa evolves into any other psychiatric disorder, and anorexia nervosa is a very unusual outcome. Body weight changes little over time and, in contrast with anorexia nervosa, the mortality rate appears not to be raised. Robust predictors of long-term course or outcome have been identified.

Acknowledgements

We are grateful to the Wellcome Trust for its support. CGF holds a Principal Research Fellowship (046386). ZC and RM are supported by a programme grant (046386).

Further information

Fairburn, C.G. (ed.). (2008). *Cognitive behavior therapy and eating disorders.* Guilford Press, New York.

International Journal of Eating Disorders.

References

1. Russell, G.F.M. (1979). Bulimia nervosa: an ominous variant of anorexia nervosa. *Psychological Medicine*, **9**, 429–48.

2. American Psychiatric Association. (1980). *DSM-III: diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Association, Washington, DC.
3. American Psychiatric Association. (1987). *DSM-III-R*. American Psychiatric Association, Washington, DC.
4. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
5. World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders. Author, Geneva.
6. Fairburn, C.G. and Bohn, K. (2005). Eating disorder NOS (EDNOS): an example of the troubl esome 'not otherwise specified' (NOS) category in DSM-IV. *Behaviour Research and Therapy*, **43**, 691–701.
7. Fairburn, C.G. and Harrison, P.J. (2003). Eating disorders. *Lancet*, **361**, 407–16.
8. Fairburn, C.G., Cooper, Z., and Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: a 'transdiagnostic' theory and treatment. *Behaviour Research and Therapy*, **41**, 509–28.
9. Walsh, B.T. (1993). Binge eating in bulimia nervosa. In *Binge eating: nature, assessment and treatment* (eds. C.G. Fairburn and G.T. Wilson), pp. 37–49. Guilford Press, New York.
10. Stein, A., Woolley, H., Cooper, S.D., *et al.* (1994). An observational study of mothers with eating disorders and their infants. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, **35**, 733–48.
11. Stein, A., Murray, L., Cooper, P., *et al.* (1996). Infant growth in the context of maternal eating disorders and maternal depression: a comparative study. *Psychological Medicine*, **26**, 569–74.
12. Fairburn, C.G., Cooper, Z., Bohn, K., *et al.* (2007). The severity and status of eating disorder NOS: implications for DSM-V. *Behaviour Research and Therapy*, **45**, 1705–15.
13. Strober, M., Freeman, R., Lampert, C., *et al.* (2000). Controlled family study of anorexia nervosa and bulimia nervosa: evidence of shared liability and transmission of partial syndromes. *American Journal of Psychiatry*, **157**, 393–401.
14. Fairburn, C.G., Welch, S.L., Doll, H.A., *et al.* (1997). Risk factors for bulimia nervosa—a community-based case-control study. *Archives of General Psychiatry*, **54**, 509–17.
15. Lilenfeld, L.R., Kaye, W.H., Greeno, C.G., *et al.* (1998). A controlled family study of anorexia nervosa and bulimia nervosa—psychiatric disorders in first-degree relatives and effects of proband comorbidity. *Archives of General Psychiatry*, **55**, 603–10.
16. Mannucci, E., Rotella, F., Ricca, V., *et al.* (2005). Eating disorders in patients with Type 1 diabetes: a meta-analysis. *Journal of Endocrinological Investigation*, **28**, 417–9.
17. Fairburn, C.G. and Cooper, P.J. (1982). Self-induced vomiting and bulimia nervosa: an undetected problem. *British Medical Journal*, **284**, 1153–5.
18. Fairburn, C.G. and Cooper, P.J. (1984). Binge-eating, self-induced vomiting and laxative misuse—a community study. *Psychological Medicine*, **14**, 401–10.
19. Fairburn, C.G. and Beglin, S.J. (1990). Studies of the epidemiology of bulimia nervosa. *American Journal of Psychiatry*, **147**, 401–8.
20. Hoek, H.W. and van Hoeken, D. (2006). Review of the prevalence and incidence of eating disorders. *International Journal of Eating Disorders*, **34**, 383–96.
21. Keel, P.K. and Klump, K.L. (2003). Are eating disorders culture-bound syndromes? Implications for conceptualizing their etiology. *Psychological Bulletin*, **129**, 747–69.
22. Kendler, K.S., MacLean, C., Neale, M., *et al.* (1991). The genetic epidemiology of bulimia nervosa. *American Journal of Psychiatry*, **148**, 1627–37.
23. Fairburn, C.G., Cooper, Z., Doll, H.A., *et al.* (1999). Risk factors for anorexia nervosa—three integrated case-control comparisons. *Archives of General Psychiatry*, **56**, 468–76.
24. Slof-Op't Landt, M., van Furth, E.F., Meulenbelt, I., *et al.* (2005). Eating disorders: from twin studies to candidate genes and beyond. *Twin Research and Human Genetics*, **8**, 467–82.
25. Bulik, C.M., Sullivan, P.F., Wade, T.D., *et al.* (2000). Twin studies of eating disorders: a review. *The International Journal of Eating Disorders*, **27**, 1–20.
26. Kaye, W.H., Frank, G.K., Bailer, U.F., *et al.* (2005). Serotonin alterations, in anorexia and bulimia nervosa: new insights from imaging studies. *Physiology & Behavior*, **85**, 73–81.
27. Kaye, W.H., Wagner, A., Frank, G., *et al.* (2006). Review of brain imaging in anorexia and bulimia nervosa. In *Annual review of eating disorders Part 2* (eds. S. Wonderlich, J.E. Mitchel, M. de Zwaan, and H. Steiger), pp. 113–29. Radcliffe, Oxford.
28. Fairburn, C.G. and Cooper, Z. (1993). The eating disorder examination (12th edn). In *Binge eating: nature, assessment and treatment* (eds. C.G. Fairburn and G.T. Wilson), pp. 317–60. Guilford Press, New York.
29. Garner, D.M. (1991). *Eating disorder inventory-2*. Psychological Assessment Resources, Odessa, FL.
30. Fairburn, C.G. and Beglin, S.J. (1994). Assessment of eating disorders: interview or self-report questionnaire? *The International Journal of Eating Disorders*, **16**, 363–70.
31. National Collaborating Centre for Mental Health. (2004). Eating disorders: core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders. British Psychological Society and Royal College of Psychiatrists, London.
32. Wilson, G.T. and Fairburn, C.G. (2007). Treatments for eating disorders. In *A guide to treatments that work* (3rd edn) (eds. P.E. Nathan and J.M. Gorman). Oxford University Press, New York.
33. Romano, S.J., Halmi, K.A., Sarkar, N.P., *et al.* (2002). A placebo-controlled study of fluoxetine in continued treatment of bulimia nervosa after successful acute fluoxetine treatment. *American Journal of Psychiatry*, **159**, 96–102.
34. Fluoxetine Bulimia Nervosa Collaborative Study Group. (1992). Fluoxetine in the treatment of bulimia nervosa. A multicenter, placebo-controlled, double blind trial. *Archives of General Psychiatry*, **49**, 139–47.
35. Fairburn, C. (1981). A cognitive behavioural approach to the treatment of bulimia. *Psychological Medicine*, **11**, 707–11.
36. Fairburn, C.G., Marcus, M.D., and Wilson, G.T. (1993). Cognitive-behavioral therapy for binge eating and bulimia nervosa: a comprehensive treatment manual. In *Binge eating: nature, assessment and treatment* (eds. C.G. Fairburn and G.T. Wilson), pp. 361–404. Guilford Press, New York.
37. Agras, W.S., Crow, S.J., Halmi, K.A., *et al.* (2000). Outcome predictors for the cognitive behavior treatment of bulimia nervosa: data from a multisite study. *American Journal of Psychiatry*, **157**, 1302–8.
38. Fairburn, C.G., Agras, W.S., Walsh, B.T., *et al.* (2004). Prediction of outcome in bulimia nervosa by early change in treatment. *American Journal of Psychiatry*, **161**, 2322–4.
39. Wilson, G.T., Fairburn, C.C., Agras, W.S., *et al.* (2002). Cognitive-behavioral therapy for bulimia nervosa: time course and mechanisms of change. *Journal of Consulting and Clinical Psychology*, **70**, 267–74.
40. Fairburn, C.G., Jones, R., Peveler, R.C., *et al.* (1993). Psychotherapy and bulimia nervosa: longer-term effects of interpersonal psychotherapy, behavior therapy, and cognitive-behavior therapy. *Archives of General Psychiatry*, **50**, 419–28.
41. Klerman, G.L., Weissman, M.M., Rounsaville, B.J., *et al.* (1984). *Interpersonal psychotherapy of depression*. Basic Books, New York.
42. Fairburn, C.G. (1997). Interpersonal psychotherapy for bulimia nervosa. In *Handbook of treatment for eating disorders* (eds. D.M. Garner and P.E. Garfinkel), pp. 278–94. Guilford Press, New York.

43. Agras, W.S., Walsh, B.T., Fairburn, C.G., *et al.* (2000). A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *Archives of General Psychiatry*, **57**, 459–66.
44. Fairburn, C.G. (1995). *Overcoming binge eating*. Guilford Press, New York.
45. Cooper, P.J. (1995). Bulimia nervosa and binge eating: a guide to recovery. Robinson, London.
46. Schmidt, U.H. and Treasure, J.L. (1993). *Getting better bit(e) by bit(e)*. Erlbaum, London.
47. Mitchell, J.E., Halmi, K., Wilson, G.T., *et al.* (2002). A randomized secondary treatment study of women with bulimia nervosa who fail to respond to CBT. *The International Journal of Eating Disorders*, **32**, 271–81.
48. Fairburn, C.G. (ed.) (2008). *Cognitive behavior therapy and eating disorders*. Guilford Press, New York.
49. Fairburn, C.G., Cooper, Z., Doll, H.A., *et al.* (2000). The natural course of bulimia nervosa and binge eating disorder in young women. *Archives of General Psychiatry*, **57**, 659–65.
50. Keel, P.K., Mitchell, J.E., Miller, K.B., *et al.* (1999). Long-term outcome of bulimia nervosa. *Archives of General Psychiatry*, **56**, 63–9.

4.11

Sexuality, gender identity, and their disorders

Contents

4.11.1 Normal sexual function

Roy J. Levin

Introduction

Normal sexual function means different things to different people. It is studied by a variety of disciplines: biology, physiology, psychology, medicine (in the domains of endocrinology, gynaecology, neurology, psychiatry, urology, and venereology), sociology, ethology, culture, philosophy, psychoanalysis, and history. There is often little liaison or cross-fertilization between these disciplines and each has its own literature and terminology. Some are regarded as 'hard science', suggesting hypotheses that can be supported or rejected by experiment, observation, or measurement (evidence-based). Others are looked on as 'soft science', where individual and anecdotal evidence are the norm and are encouraged.

As space is limited, this chapter will characterize 'normal sexual activity' in the Western world mainly from biological, physiological, and psychological aspects but will occasionally utilize other disciplines when they yield insights not available from the 'harder sciences'.

Biological determinants of normal sexual function

Humans are the highest evolved primates. A number of our anatomical/biological features unrelated to reproduction have been described as strongly enhancing our sexual behaviour when compared with other primates,[1] although recent studies have shown that the bonobos (pygmy chimpanzees) also use sex for reasons unconnected with reproduction.[2]

In brief, these features are as follows:

1 The relative hairlessness of our bodies allows well-defined visual displays (see point 6 below) and enhanced tactile skin sensitivity.

2 The clitoris, which is an organ whose sole function is for inducing female sexual arousal/pleasure.

3 Orgasms, in both male and female, provide intense euphoric rewards for undertaking sexual arousal to completion. The female is able to have multiple serial orgasms.

4 The largest penis among primates, whether flaccid or erect, the latter acting as a good sexual stimulator of the female genitalia.

5 Concealed (cryptic) ovulation which could influence males to undertake coitus more frequently to create pregnancy and prevent cuckolding.

6 Well-defined visual sexual displays in the female that are not linked to season or fertility (i.e. breasts, pubic hair, buttocks, and lips). The everted mucous membranes of the lips serve both as a surface display (red, moist, and shiny), and for haptic stimulation during kissing and sucking.

7 Ability of the female to undertake sexual arousal and coitus independent of season, hormonal status, or ovulation. Human females (unlike other primates) can and often do willingly partake of sexual activity and coitus when they are menstruating, pregnant, or menopausal.

8 Development of large mammary glands during puberty which act as visual sexual signals in most cultures.

These biological determinants are augmented by socio-cultural factors:

1 Language, art, and music for erotic stimulation.

2 Facial adornment with make-up to heighten appearance and sex displays (viz., lipstick).

3 Clothing, especially of the female, such as brassières to redefine the shape of breasts, corsets to redefine the shape of the body, and high-heeled shoes to elongate the legs and thrust out the

buttocks. Young adult males use tight trousers to create a genital 'bulge' and to emphasize firm rounded buttocks, the latter being a highly sexually attractive feature to young women.

4 Perfumes and scents to enhance body aroma.

The last three features (2, 3, and 4) use artificial means to enhance normal sexual signals.[1] These biological and socio-cultural factors give human sexual activity an increased appetitiveness and make it more rewarding.

Sexuality as a social construct and the concept of sexual scripting

Laqueur[3] suggests that while the sexual biology remained unchanged, its expression has been influenced over the centuries by culture, social class, ethnic group, and religion. This concept, that human sexuality is a social construct, has been strongly argued by Foucault[4] and promoted by other social constructionist authors.[5] Gagnon and Simon[6] introduced the concept of 'sexual scripting'. Scripts organize and determine the circumstances under which sexual activity occurs, they are involved in 'learning the meaning of internal states, organizing the sequences of specific sexual acts, decoding novel situations, setting limits on sexual responses and linking meanings from non-sexual aspects of life to specifically sexual experience'. Money[7] employed a similar construction in his development of 'love maps' for the individual. While patterns of behaviour are influenced by society and social forces, there is a dearth of evidence to show that sexual identity, orientation, or sexual mechanisms are also influenced.

Modelling normal sexual function—(i) the sex survey

One obvious way of describing normal sexual function is to ask people what they do. Two classic sex surveys were conducted by Kinsey and his coworkers who reported the results of interviews with 12 000 males in 1948[8] and 8000 females in 1953.[9] Their technique of sampling was to interview everyone in specific cooperating groups (clubs, hospital staff, universities, police force, school teachers, etc.). This gave samples of convenience but not a valid statistical sampling of the population. Despite their age and faulty sampling, however, there are still useful data in these surveys. In the sexual climate of the 1950s many of the findings were regarded as highly controversial. Clement[10] has reviewed the subsequent studies of human heterosexual behaviour up to 1990.

Surveys give a selective picture of sexual function. Results depend on the formulation of the questions, they rely on self-reports, and they represent only those prepared to describe their sexual behaviour. It is known, for example, that females tend to under-report their premarital sexual experiences[11] while males tend to over-report their lifetime partners.[11] Berk *et al.*[12] studied the recall by 217 university students of their sexual activity over a 2-week period assessed by questionnaires answered 2 weeks after the recording period, and by daily diaries kept over the same 2 weeks. Subjects reported more sexual activity in the questionnaires than in their diaries. Women reported giving and having more oral sex than the men. Clearly, data from questionnaire surveys should be treated cautiously.

A survey tells only what is frequent and not necessarily what is normal, but the most frequent practices often become identified with normal sexual behaviour. Surveys also vary in the range of behaviours that are asked about, for example coitus without condoms is important in the age of AIDS. Surveys have one great disadvantage, the facts that they produce are often 'perishable'; many aspects of the sex surveys of the pre-pill era, or more recently the pre-AIDS era, are now of use only in a historical or comparative basis.

Two recent well-organized surveys based on samples of the whole population have been undertaken, one in the United States and the other in the United Kingdom. Interestingly, in both surveys, questions about masturbation were disliked by the respondents. In the American survey these questions were asked in a separate self-administered questionnaire, while in the British survey they were abandoned.

The American survey[13] was conducted face to face with 3159 selected individuals who spoke English in representative households by 220 trained interviewers (mainly women). Nearly 80 per cent of the individuals chosen agreed to be interviewed. Men thought about sex often, more than 50 per cent having erotic thoughts several times a day, while females thought about sex from a few times a week to a few times a month. The frequency of partnered sex had little to do with race, religion, or education. Only three factors mattered: age, whether married or cohabiting, and how long the couple had been together. Fourteen per cent of males reported having no sex in the previous year, 16 per cent had sex a few times in the year, 40 per cent a few times a month, 26 per cent two to three times a week, and 8 per cent four times a week. The percentages were similar for women. The youngest and the oldest people had the least sex with a partner; those in their 20s had the most. Of the women aged 18 to 59 years, approximately one in three said they were uninterested in sex, and one woman in five said sex gave her no pleasure. Unlike frequency, reported sexual practices do depend on race and social class. Most practices other than vaginal coitus were not very attractive to the vast majority. In women aged 18 to 44 years of age, 80 per cent rated vaginal coitus as 'very appealing' and an additional 18 per cent rated it as 'somewhat appealing'. Among men 85 per cent regarded vaginal coitus as 'very appealing'. The most appealing activity second after coitus was watching the partner undress, and this was appealing to more men (50 per cent) than women (30 per cent). This reflects the greater voyeuristic nature of men and their willingness to pay to look at women undressing or undressed.

In regard to oral sex, both men and women liked receiving more than giving. This practice varied markedly with race and education, with higher reported rates among better educated white people than among less educated and black people. Some 68 per cent of all women had given oral sex to their partner and 19 per cent experienced active oral sex the last time they had intercourse. Seventy-three per cent of all women had received oral sex from the partner, and 20 per cent had received it the last time they had had intercourse. Corresponding experiences were reported by men.

This survey, unlike many earlier ones, asked about anal sex. Of females aged 18 to 44, 87 per cent thought it not at all appealing, and only 1 to 4 per cent thought it very or somewhat appealing. In males of the same age 73 per cent thought it not at all appealing and rather more than women thought it very or somewhat appealing. Similar reports were obtained from women and men aged 44 to 59.

Regarding masturbation, older people (over 54 years old) had lower rates than at any other age, indicating that they do not use masturbation to compensate for an overall decrease in sexual activity with their partners.

In the United Kingdom survey,[14] 18 876 people were interviewed by 488 interviewers (of whom 421 were women). The sampling used one person per address and the acceptance rate was 71.5 per cent. Questions were asked about the frequency of vaginal coitus, oral sex, and anal sex, but not masturbation. The median number of occasions of sex with a man or woman was five times a month for females aged 20 to 29 and males aged 25 to 34, but declined to a median of two per month for males aged 55 to 59. More than 50 per cent of the females in the 55 to 59 age group reported no sex in the last month, but in this age group females are more likely than men to have no regular partner because they are widowed, separated, or divorced.

Vaginal coitus was reported by nearly all females and males by the age of 25. Fifty-six per cent of males and 57 per cent of females reported vaginal coitus in the previous week, and non-penetrative sex was practiced by 75 per cent men and 82 per cent of women. Twenty-five per cent of males had genital stimulation in the previous 7 days. Cunnilingus and fellatio were common but less practiced than vaginal coitus. Of men and women aged 18 to 44, 60 per cent had oral sex in the previous year but in the 45- to 59-year-old group this fell to 30 per cent for women and 42 per cent for men. This and other sex surveys suggest that the practice of oral sex has increased since the 1950s and 1960s. Anal coitus was infrequent; approximately 14 per cent of the males and 13 per cent of the females had ever undertaken it, and only 7 per cent of males or females had practiced it in the previous year.

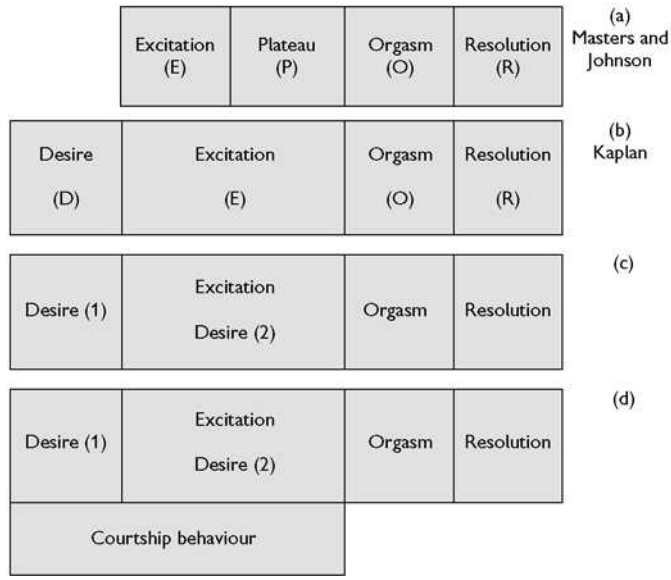


Fig. 4.11.1.1 The development of the human sexual response model from (a) the original **EPOR** model of Masters and Johnson[15] through (b) the desire, excitation, orgasmic, and resolution (**DEOR**) model of Kaplan[17] to (c) the proposed modification with desire phase 1 (before initiation of the excitation phase and desire phase 2 during excitation phase) and finally (d) with added courtship behaviour.

Modelling normal sexual function—(ii) the sexual response cycle

A direct way of investigating normal sexual function is to observe and measure the body changes that take place when men and women become sexually aroused. From these data, models have been constructed of the normal sequence of changes during sexual arousal, coitus, and orgasm. The first models described a simple sequence of increasing arousal and excitement culminating in rapid discharge by orgasm, displayed graphically as an ascent, peak, and then descent. As the investigations became more sophisticated, understanding of the body responses grew and the models became more detailed and complex.[5,15,16]

The EPOR model—a sexual response cycle model

A most successful human sexual response model was that formulated by Masters and Johnson.[15] In the laboratory, they observed the changes that took place in the male and female body and especially the genitals during sexual arousal to orgasm either by masturbation or by natural or artificial coitus with a plastic penis that allowed internal filming of the female genitalia. After studying approximately 7500 female and 2500 male arousals to orgasm in some 382 female and 312 male volunteers over 11 years, they proposed a four-phase, sequential, and incremental model of the human sexual response cycle (Fig. 4.11.1.1). The phases were described as the excitation (**E**) phase (stimuli from somatogenic or psychogenic sources raise sexual tensions), the plateau (**P**) phase (sexual tensions intensified), the orgasmic (**O**) phase (involuntary pleasurable climax), and finally the resolution (**R**) phase (dissipation of sexual tensions). The great success of this **EPOR** model was its wide compass; it could characterize the sexual responses of women and men, both heterosexual and homosexual, ranging from simple petting to vaginal or anal coitus with or without orgasm. However, it had several weaknesses.

Modifying the EPOR model into the DEOR model

The first weakness of the EPOR model is that it was derived from the study of a highly selected group of American men and women volunteers who could arouse themselves to orgasm in a laboratory, on demand, and allow themselves to be watched/filmed or measured for scientific and altruistic (or perhaps exhibitionistic) purposes. The second weakness was the lack of interobserver agreement about the changes observed and of confirmation of their sequential reliability. Robinson[16] examined the **E** phase and **P** phase, and concluded convincingly that the **P** phase was simply the final stage of the **E** phase. Helen Kaplan,[17] a New York sex therapist, proposed that before the **E** phase there should be a 'desire phase' (**D** phase). This proposal came from her work with women who professed to have no desire to be sexually aroused, even by their usual partners. She suggested that the desire must occur before sexual arousal can begin. Kaplan's subjects were attending a clinic and remarkably no studies were ever conducted with a control normal population (either women or men) to investigate whether this 'self-evident' fact was true. Despite this, the **EPOR** model gradually became replaced by the desire, excitation, orgasmic, and resolution phase (**DEOR**) modification. While this is the currently accepted model, the centrality of the desire phase in women remains uncertain (Fig. 4.11.1.1). In a survey of non-clinic sexually experienced women in Denmark, about a third reported that they never experienced spontaneous sexual desire[18] and in an American survey women reported periods of several months when they lacked interest in sex.[13] The other problem with the desire phase is its location in the sequential **DEOR** model.

Sexual desire (D1-proceptive desire) that appears to be spontaneous (but presumably must still be activated by a trigger) should obviously be placed at the beginning of the model (Fig. 4.11.1.1b), while sexual desire (D2-receptive desire) created when the person is sexually aroused by another occurs during the E-phase (Fig. 4.11.1.1c).

It has been proposed[19] that while the DEOR model fits for females in younger couples' relationships for longer maintained ones sexual activity is undertaken often for factors such as intimacy, security, and acceptance and becomes more influenced by cognitive and emotional processes and the possible outcomes of the experience (e.g. mutual pleasure, confirming commitment and trust, enhancing emotional intimacy) rather than proceptive/receptive desires.

Courtship (mating) behaviour—activity initiating normal sexual behaviour

With the possible exception of rape, the pre-initiation of sexual activity normally starts with flirting/courtship behaviours. Sometimes this activity can precede the desire phases; sometimes it occurs during the desire phases, and sometimes in the excitation phase (Fig. 4.11.1.1d).

Evolutionary psychology attempts to explain the strategies of mating.[20] Its message is not always palatable to modern sensitivities about sexual equality. It is argued that women invest more in their offspring than men,[21] that this investment is a scarce resource that men compete for, and that men can enhance their reproductive strategy by mating frequently. Most men are first visually attracted to a possible female sexual partner. They look for youthfulness and physical attractiveness in the form of regular features (symmetry), smooth complexion, optimum stature, and good physique, and they value virginity and chastity. Partner variety is highly desired. Women, however, need to obtain high-quality mates with abundant resources and look for emotional and financial status and security. Clearly the strategies conflict giving rise to different preferences in mate choice and casual sex, and different levels of investment or commitment to relationships.[22]

Once the chosen female (or male) accepts the initiation of flirting/courtship behaviour, the subsequent stages form a stereotyped sequence which is found in many different cultures. The stages are look, approach, talk, touch, synchronize, kiss (caress), sex play, coitus. It is a sequence that the poet Ovid knew in the first century BCE. Morris[1] characterized human courtship behaviour further into 12 basic stages: eye to body, eye to eye, voice to voice, hand to hand, arm to shoulder, arm to waist, mouth to mouth, hand to head, hand to body, mouth to breast, hand to genitals, genitals to genitals. Similar hierarchies have been constructed extending the behaviour to oral-genital contacts.

Although kissing has been described as 'an inhibited rehearsal for intercourse and other sexual practices'[23] and is usually undertaken in the courtship behaviour well before genital activity occurs, it is sometimes thought of as more intimate than coitus. Prostitutes, for example, traditionally do not kiss their clients on the mouth, reserving the activity for their private sexual behaviour. Nicholson[24] has speculated that kissing may be a mechanism by which semiochemicals (similar to pheromones) are exchanged between humans to induce bonding.

Common extragenital changes during sexual arousal

In both males and females, effective sexually arousing stimuli cause a number of physiological changes.[15,25] There is increased respiration, heart rate, and blood pressure, nipple erection, often sweating, and a sex flush (maculopapular skin rash). Muscle tension increases (myotonia) and the pupils dilate. All these changes become more intense as arousal increases. Following orgasm the changes dissipate rapidly (R phase); without orgasm they dissipate more slowly.

The endocrinology of normal sexual function

(a) Peptide hormones and neuropeptides

Prolactin, a peptide hormone, is secreted by cells of the anterior pituitary in both males and females. In females it is a key lactogenic hormone involved in stimulating the manufacture of breast milk but it has no proven physiological reproductive function in males. Pathologically high plasma levels of prolactin (hyperprolactinaemia) are accompanied by disturbances of sexual/reproductive function especially erectile dysfunction in males and inhibition of ovulation in females. The exact causes are unknown. As it is secreted in higher amounts at orgasm in both sexes it has been claimed to be a major factor responsible for 'switching off' sexual arousal but this ignores the fact that women are multiorgasmic while the evidence for this function even in males is unconvincing.[26]

Two neuropeptides, oxytocin and vasopressin (Antidiuretic Hormone, ADH) are hormones and also act as neuropeptide transmitters with distribution in parts of the brain and spinal cord. They are secreted as hormones by the hypothalamus/posterior pituitary with increases at orgasm in both males and females. Oxytocin is often claimed to be responsible for facilitating smooth muscle contractions during ejaculation and uterine contractions during orgasm but the possible role of vasopressin has been ignored. In some animals, oxytocin is involved in pair bonding but the evidence for this in humans is poor.

Males

(a) Androgens

The major steroids influencing normal sexual function in males are the androgens secreted by the testes mainly as testosterone with much smaller quantities of androstenedione and dihydrotestoterone. The adrenal cortex also manufactures and secretes androgens but this amounts only to 2 per cent of the total. Of the total testosterone (260–1000 ng/dl), 95 per cent is bound to plasma proteins; only the unbound fraction (34–194 pg/ml) is an active virilizing hormone.

The development and maintenance of the masculine musculature, bone growth, genitals, and pubic and axillary hair is androgen dependent. The mechanism of the hormonal masculinization of the brain in rodents involves the aromatization of testosterone, but in humans the role of aromatization is still uncertain.

(b) Androgens and male sexual behaviour

The involvement of androgens in adult human male sexual behaviour has been reviewed many times (see Levin[27] for references). Removal of testosterone by castration usually leads to a decrease in sexual activity and drive (libido) in the majority of subjects within

12 months. There is, however, large individual variations, and some castrates retain sexual activity and interest for years.[28] Factors such as adrenal androgens, the availability of sexual partners, and attitude to the operation influence the response. In castrates and in hypogonadal males, replacement of testosterone restores sexual interest and activity.

Females

(a) Oestrogen, progesterone, and androgens

In the female, oestrogen, progesterone, and androgens are involved in differentiating and maintaining genital and breast tissues and in influencing normal sexual function. Female genital development, unlike that of the male, does not appear to need hormonal stimulation as the development of female genital structures are the foetal prototype or default programme.[29] During puberty, ovarian oestrogen and progesterone together with androgens from the adrenal cortex induce growth and functional changes in the internal and external genitalia, breast, and nipples. Oestrogens induce growth in the fallopian tubes, uterus, vagina, and breasts and lay down subcutaneous fat largely in the breast, hip, and thigh regions to create the rounded contours of the female body that are highly attractive to the male. The fat laid down is enough to supply the energy for a pregnancy and the subsequent lactation. Females have a plasma androgen concentration some 10 times less (15–70 ng/dl) than those in the male. Androgens are produced by the ovaries (25 per cent), adrenals (25 per cent) and in peripheral tissues from adrenal-secreted androgen precursors. They are responsible for the development and maintenance of the clitoris, nipples, pubic and axillary hair, and probably the labia, periurethral glans and pelvic-striated musculature. The variation of the androgen levels in the plasma during the menstrual cycle is small and it is the free androgen level (1–21 pg/ml) not the bound that is the active principle.

(b) Role of hormones in female sexuality

While over 60 studies have been undertaken to examine whether the changing hormonal levels of the menstrual cycle influence the sexual arousal of the female,[30] neither oestrogen or progesterone have been found convincingly to play a direct role in influencing the sexual activity of the human female apart from their indirect functions in the maintenance of the structures and functions of the female genitals, especially the vagina.

The role of androgens in female sexuality is not clear-cut.[31] Some propose that, as in the male, it is the major hormonal influence on the female libido. Removal of the adrenals has been shown to reduce desire and ability to reach orgasm. Excess androgens stimulates libido but in pharmacological not physiological doses. Such doses affect the structure and sensitivity of the clitoris (an androgen-sensitive tissue), so the effects on sexuality might not be only brain mediated.

(c) Sexual behaviour during the menstrual cycle

Despite numerous studies it is still uncertain whether female sexual behaviour is influenced by the hormonal changes in the menstrual cycle. Meuwissen and Over[30] surveyed 64 published studies. A significant number showed a premenstrual peak in sexual desire and activity, others a postmenstrual peak, either at menstruation or ovulation but the latter studies often used poor methodology to determine the time of ovulation.

Male genital functions during normal sexual arousal

While the **DEOR** model characterizes the general sexual arousal of humans, a more specific detailed physiological model in males is that of excitation, erection, emission, and ejaculation with orgasm. Each of these is served by separate mechanisms. Although ejaculation and orgasm usually occur temporally together, they also have separate mechanisms.

(a) Excitation

Sexual excitation can occur through any of the five senses by psychogenic or somatogenic stimuli. In special circumstances arousal can become linked to or greatly enhanced by fetishitic association with non-sexual objects such as feminine garments of underwear, rubberware, shoes, furs, etc.

The sexual excitation is manifested in the brain by activation of numerous areas (see section on brain activation during human sexual arousal). Centres in the spinal cord for erection and ejaculation are known from animal studies and it is likely that they also exist in the human cord. They are activated by neural efferent activity from the aroused areas of the brain and initiate erection and ejaculation.

(b) Erection: the conversion of the flaccid urinary penis to the rigid sexual penis

The three longitudinal erectile chambers of the penis are arranged with a side-by-side dorsal pair of corpora cavernosa above the single ventral corpus spongiosum. The corpora cavernosa are covered by the tunica albuginea, a 2-mm thick fibrous membrane which is resistant to stretch. The corpus spongiosum surrounds the length of the penile urethra and is enlarged at its base to form the urethral bulb and distally to create the glans penis. While it becomes engorged with blood during arousal it is not involved in the rigidity of the erection but protects the urethra from closure. The unaroused penis is flaccid because the pudendal arterial blood flow into the erectile tissues is limited by the high sympathetic (adrenergic-mediated) constrictive tone in the smooth muscle of the vessels of corpus cavernosum. On sexual arousal, the sympathetic tone is reduced; the neural innervation of the arteries and cavernosal chambers is activated to release vasoactive intestinal peptide (VIP), a peptidergic vasodilator neurotransmitter that directly relaxes smooth muscle and nitric oxide (NO), the nitrergic neurotransmitter. NO activates the enzyme cyclic guanylase in the smooth muscle cells of the cavernous tissue and blood-vessel endothelium to produce cyclic guanosine monophosphate (cGMP), the second messenger that creates intracellular conditions to relax the muscle. The enzyme phosphodiesterase 5 breaks down the cGMP and inhibitors of this enzyme, which can be taken by mouth, facilitate the attainment of erection even with reduced neural inputs.

The vasodilatation of the arterial supply by VIP together with the relaxation of the vessels of the cavernosal tissue allows them to fill under arterial pressure stretching the chambers until they become stiff against their covering of unyielding tunica albuginea, and the veins (emissary) that pass obliquely through the tunica become occluded greatly reducing penile vascular drainage. The flaccid urinary penis has been converted into the erect rigid sexual penis some 7 to 8 cm longer. The rigidity is essential for successful vaginal penetration and to stimulate its walls (especially the anterior) during penile thrusting. The striated muscles of the pelvic region, namely the ischiocavernosus and bulbocavernosus, are not

normally involved in creating penile erection[32] although they can be voluntarily contracted in short bursts to aid its rigidity. The engorged corpus spongiosum is less rigid than the cavernosal chambers making the glans of the penis softer and less damaging to the female labia and vagina.

(c) Internal genitals

The genital fluids of the testes, epididymis, and accessory genital glands of the male are involved in emission. These glands are the bulbo-urethral (Cowper's gland), the prostate (approximately 30 per cent of the total volume of the ejaculate), and the paired seminal vesicles (approximately 60 per cent of the volume of the ejaculate). The fluids from all these together with that of the glands of Littré that line the penile urethra, constitute the ejaculate or semen which has a characteristic odour and rapidly forms a coagulum in contact with air. Subsequently enzymes in the semen break down the coagulum and release the trapped sperm.

(d) Emission

This phase begins with the movement of the various genital fluids into the ducts initiated by the neurally induced contraction of smooth muscles in the capsules of the testes, epididymis, and seminal vesicles. The secretions spurt into the prostatic urethra, and the sphincter of the bladder neck closes to prevent reflux into the bladder. When this happens the male experiences the sensation of 'ejaculatory inevitability' and knows that he will ejaculate within a second or two and that conscious suppression of the ejaculatory reflex is now impossible. The contractions of the smooth muscle of the glandular capsules together with the contraction of the vas deferens and peristalsis in the urethra move the semen along into the penile urethra.

(e) Ejaculation

Within a second or two later the bulbocavernosus muscle of the perineal region contracts clonically, initially at about 1 per 0.8 s, squeezing the urethra and forcing out the ejaculate. As ejaculation proceeds, the interval between each striated muscle contraction gets longer and their force weaker until they gradually die out.[15,32] Their number can vary between 5 and 60. Most of the ejaculate is expressed within the first half dozen contractions. If the striated muscles are paralysed the semen is squeezed out only by the smooth muscle peristaltic contractions which produce a dribbling ejaculate with no projectile force and little pleasurable quality.

(f) Male orgasm

Orgasm, the supreme ecstatic pleasure is experienced just before the striated contractions occur and is then associated, throbbingly, with each subsequent contraction slowly decreasing in intensity and dying away as do the contractions. It is felt as an intense pleasurable throbbing/pumping in the penis and pelvic area and can last from 5 to 60 s. Most males groan with each squirting contraction. Kinsey *et al.*[8] marshalled the evidence showing that orgasm and ejaculation were separate mechanisms. Briefly, orgasm occurs without ejaculation in preadolescent males, in some adult males orgasm does not occur until a few seconds after ejaculation, a few adult males are anatomically incapable of ejaculation but have orgasms, and males who have been prostatectomized cannot have ejaculations but some can have orgasms.

The intensity of orgasm usually varies with the duration of the sexual arousal (the longer it is maintained the greater the subsequent orgasm), the erotic excitement and novelty of the arousing stimuli, and previous ejaculation, especially the interval from the last one (initial ejaculations have usually better orgasms than subsequent ones). Males have a post ejaculation refractory time (PERT) and usually cannot have an erection or another ejaculation until some time has passed. The PERT varies with age and can be anything from minutes, when young, to hours or days when older.[15] It is not known where this inhibitory mechanism resides but animal work suggests that it is in the brain rather than the spinal cord.[33] Some men claim to be able to learn to inhibit ejaculation and yet have repeated serial orgasms.[34]

It has been stated that the larger the ejaculate volume the greater is the orgasmic pleasure,[15] the studies however were flawed because they used men who increased semen volume by abstaining from ejaculation for days. This confounds the effects of semen volume with that of ejaculatory abstinence which itself enhances subsequent sexual pleasure. In fact semen volume does not appear to be the arbiter of pleasure or the trigger for ejaculation.[35] Drugs can induce a 'dry ejaculation' but the pleasure of the orgasm appears unimpaired,[33] and in young boys the pleasure of the early dry orgasm is not noticeably changed when semen becomes added to the ejaculation around puberty.[36]

An ignored feature is whether there is a typology of male orgasms. Most orgasms arise from penile stimulation but they can also be activated by per rectum digital stimulation of the prostate gland. No laboratory study of these orgasms has ever been made but anecdotal reports say that they feel deeper, more widespread and intense and last longer than those from penile stimulation.[37]

Female genital functions during normal sexual arousal

(a) External

(i) Labia

The external female genitalia consist of the outer (majora) and inner (minora) labia containing erectile tissue that surround the vaginal introitus. Normally the outer labia meet and cover the introitus, but in some women the inner labia protrude even when they are sexually unaroused. Sexual arousal creates vasocongestion especially in the labia minora which protrude through the majora adding approximately 1 to 2 cm to the length of the vagina. The labia minora become erotically sensitive to touch and friction when engorged.

(ii) Clitoris

Although the clitoris is the homologue of the penis, its precise anatomical structure is still uncertain. The most recent description by O'Connell *et al.*[38] is of a triplanar complex of erectile tissue with a midline shaft lying in the medial sagittal plane about 2 to 4 cm long and 1 to 2 cm wide, which bifurcates internally into paired curved crura 5 to 9 cm long, and externally is capped with a glans about 20 to 30 mm long with a similar diameter. Two vaginal bulbs of erectile tissue are closely applied on either side of the urethra. The shaft's erectile tissue consists of two corpora cavernosa surrounded by a fibrous sheath (tunica albuginea) and the whole is covered by a clitoral hood formed in part by the fusing of the two labia minora. The uncertainty concerns the location and extent of the female corpus spongiosum. Some describe it as wrapped around the urethra, others state that the vaginal (vestibular) bulbs on either side of the vaginal wall are spongiosus tissue and unite ventrally to the urethral meatus to form a thin strand of erectile tissue ending

in the glans. Recent studies by van Turnhout *et al.*[39] have clarified the situation. They confirmed by dissections in fresh cadavers that the bilateral vestibular bulbs terminate into the glans clitoridis.

With sexual arousal, the blood flow to the clitoris is increased probably by a mechanism involving its vipergic (VIP) and nitrinergic (NO) innervation leading to its tumescence (swelling) but, contrary to many inaccurate descriptions, without true erection (i.e. without rigidity). The enhancement of its blood flow is paralleled by an increased sensitivity to touch and friction especially of the glans.

(iii) Periurethral glans

There is a triangular area of mucous membrane that surrounds the urethral meatus, extending from just below the glans of the clitoris to the entrance of the vagina. This area has been called the 'periurethral glans'[25] and is thought to be erotically stimulated, especially during penile thrusting, as it is known that the tissue is pushed and pulled into and out of the vagina by the penile movements.[25] The periurethral glans is actually part of the corpus spongiosum if we accept the anatomical designation of van Turnhout *et al.*[39] which suggest that it is the homologue of the male glans. The extent, mobility, density of innervation, and sensitivity of this erotic site may explain why some women find it easy to reach orgasm during penile thrusting alone.

(b) Internal

(i) Vagina

No single structure can describe the vaginal shape. In sexual quiescence it is a potential space with an H-shaped cross-section and an elongated S-shaped longitudinal section culminating in a cul-de-sac, the anterior wall of which is penetrated by the cervix. The anterior and posterior walls touch but their film of basal vaginal fluid prevents adhesion. This basal fluid is a mixture of fluids from the vagina itself (basal transudate) with uterine and cervical secretions. The squamous epithelial lining of the vagina actively transports sodium ions from the vaginal fluid back into the blood. As fluid follows this ion movement osmotically the vagina is continually producing and reabsorbing its own fluid[25,40] normally lower in sodium and higher in potassium compared to plasma.

Because sexual activity is intermittent, the basal vaginal blood flow is maintained normally at a minimal level by a high adrenergic-mediated vasoconstrictive tone and vasomotion. The latter is the mechanism by which tissue viability is maintained at a basal level by not having all the capillaries open at the same time but rather each one opens and closes randomly to supply by demand its surrounding tissue needs of oxygen and metabolic substrates. On sexual arousal, the blood supply to the vaginal walls is increased by the liberation of VIP from the vipergic neural innervation. This increases the flow through the open capillaries and recruits new ones until all are open and vasomotion disappears,[41] Within seconds this creates blood-vessel engorgement and an increased plasma transudate filters out of the capillaries and percolates between the cells of the epithelium saturating its limited reabsorptive capacity. The newly formed neurogenic vaginal transudate, with its higher sodium ion concentration, creates a lubricating film on the vaginal surface which is essential for painless penile penetration and thrusting. Poor or inadequate lubrication can lead to dyspareunia (painful coitus) and subsequent sexual dysfunction. On cessation of sexual arousal or after orgasm the blood flow returns to the basal level, the fluid is reabsorbed back into the blood following the continuous absorption of sodium ions by the vaginal epithelium and vasomotion is restored.

(ii) Coitus and the vagina

The cul-de-sac of the vagina is expanded during sexual arousal and the uterus with cervix is lifted clear of its posterior wall (vaginal tenting).[15] This vaginal tenting is an important reproductive feature that delays the transport of sperm allowing the ejaculate to decoagulate and initiate motility and pre-capacitation changes that facilitate the process of sperm capacitation (see Levin[40,41] for references). In the ventral–ventral ('missionary') coital position, penile penetration and thrusting stretches and stimulates the structures of the anterior vaginal wall which include the urethra, the 'G spot' (see below), and neural structures in Halban's fascia. All these (the anterior wall erotic complex) are thought to be capable of creating erotic excitement when so stimulated, giving rise to a significant part of the sexual pleasure normally experienced by most women during coitus.

(iii) The erotic structures of the anterior vaginal wall: urethra, 'G spot', and Halban's fascia

The urethra, approximately 4 cm long, is invested with erectile tissue which becomes engorged on sexual arousal. Ultrasound imaging during coitus has shown that the thrusting penis stretches the urethra.[42] There is an area on the anterior vaginal wall a few centimeters from the introitus, at or around the junction of the urethra with the bladder, that becomes swollen and on strong pressure stimulation can induce orgasm. This urethral area was first identified by Grafenberg[43] but the observation was overlooked until its rediscovery by Perry and Whipple[44] who named it the 'G spot' in recognition of the original discoverer. Anatomically, it probably represents the 'paraurethral' or 'periurethral' glands now referred to as the 'female prostate'. In some women these produce at orgasm a small amount of fluid secretion loosely referred to as the 'female ejaculate' (see[25,44] for references). These glands are in the space between the bladder trigone and the neck of the urethra which is filled with fibroelastic mesenchymal lamina rich in vascular lacunae and contains nerve fibres and pseudocorpuscular terminals. This area, known as Halban's fascia[45] when stimulated by pressure (penile or digital) creates intense sexual pleasurable feelings.[44–46] As any pressure stimulus on the anterior vaginal wall will in fact stimulate the G spot, Halban's fascia and the urethra at the same time it makes it difficult to apportion the generation of sexual pleasure to any specific structure.

The involvement of the spinal cord in activating the female erotic structures during sexual arousal has been poorly studied and little is known of possible mechanisms.

(c) Female orgasm

As in the male, the female orgasm creates supreme ecstatic pleasure usually accompanied by throbbing striated muscle contractions, especially of the ischiocavernous and bulbospongiosus muscles, but other pelvic-striated muscles can also be involved.[15,37] Most females groan or moan during these contractions. The induction of orgasm in women by coital stimulation alone is not as frequent as that in the male, about half not achieving orgasm unless clitoral stimulation is also used. The reason for this difference is usually ascribed either to the greater inhibitory education about sexual pleasure experienced by women or the lack of correct genital

stimulation. A major difference between males and females is that females can be multiorgasmic because they do not have a PERT after orgasm.[15,23,33] There is mounting evidence that, unlike the unitary concept of the **DEOR** model, the erotic stimulation of different genital sites (especially the anterior vaginal wall compared with the clitoris) induces different types of orgasmic response both subjectively and physiologically.[44,47,48]

Brain activation during human sexual arousal

Before the advent of brain imaging by functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) to identify which areas are activated by and during sexual arousal and orgasm, only limited information was obtained by observing the effects of brain lesions on behaviour, epileptic case studies, electroencephalography (EEG) activity, and rare electrical recordings and stimulations of specific brain areas.[27] Inferences about human sexual arousal mechanisms often had to be made from animal experiments but the problem of species differences always was the spectre in the wings (see comment below).

One type of brain imaging (blood oxygen level dependent [BOLD]) relies on the concept that activation of neurones requires an increased demand for oxygen which normally entails an increase in their blood supply bringing oxygenated blood. The change in the magnetic susceptibilities of the oxygenated and deoxygenated blood is measured and is an index of the increase or decrease in flow to the neuronal area. In the PET technique a short-lived radioactive tracer (often ^{15}O) is injected intravenously and its concentration in the brain area of high blood flow localized by the radioactivity. Both techniques rely heavily on extensive computer programs to correct for a variety of essential artefact corrections one of the most important being corrections for movement. The usual protocol underlying most studies of sexual arousal is to first measure the activity of the brain area under study during a non-sexual basal state (looking at neutral videos) then switching to viewing sexually arousing videos. By subtracting the activity of the neutral from the aroused measurements the remaining activity is assumed to be due the sexual arousal per se. One difficulty is what level (threshold) of activity of a site is to be regarded as physiological. Investigators tend to choose their own criteria making direct comparisons between different studies difficult. Another difficulty is that many areas of the brain are multifunctional (e.g. the amygdala, the periaqueductal grey, the cerebellum) and are thus activated/deactivated by different stimuli (viz., pain, pleasure, fear, anger, emotional processing) the activation/deactivation may be an epiphenomenon of the arousal rather than its cause. Finally, it takes a few seconds to build up a brain scan but neuronal activation takes just hundredths of a second; the image will always lag behind the actual activation so that identifying what is the primary initiator of brain arousal will not be apparent.

Despite all the above caveats about brain imaging one outstanding conclusion is now clear from all the various studies—that there does not appear to be a single site creating arousal or orgasm, multiple site co-activation (also referred to as a neural network) is the rule in both males and females. While a number of the features of brain activation appear common to sexual arousal in the brains of both sexes, the amygdala and hypothalamus are said to be more strongly activated in men when viewing identical sexual stimuli. According to Holstege's group,[47] the first to map brain areas active during ejaculation/orgasm, the strongest activated primary region during male erection/orgasm is the ventral tegmental field region, the midbrain lateral central tegmental field, the zona incerta, the suprafascicular nucleus, the ventroposterior, midline and intralaminar nucleus, and the cerebellum. Decreased activity was seen in the amygdala and adjacent entorhinal cortex. This decrease in the amygdala is said not to occur during women's arousal.[48] Another group.[49] imaging the brain in sexually aroused males reported increased neural activity in areas that included the right frontal cortex, the inferior temporal cortex, the left anterior cingulate cortex, and the right insula. However, only a subset of these areas (anterior cingulate, insula, amygdala, hypothalamus, and secondary somatosensory cortex) were involved in creating a full erection.

Imaging studies undertaken by Komisaruk and Whipple[48] used both women who were paraplegic and some able-bodied women who had the rare ability to create orgasm by mental activity (thoughts) alone. In the former group, where arousal was induced by cervical vibratory stimulation, a large number of sites, including the hypothalamus, were activated by the stimulus used to induce orgasm but in the mental activity group, the only sites activated by thought were regions of the nucleus accumbens, the paraventricular nucleus of the hypothalamus, the hippocampus, and anterior cingulate cortex. The authors suggested that these sites may be specifically related to activate the female orgasm. Interestingly, the amygdala was not activated during these 'thought' induced orgasms. How 'normal' the 'orgasm by thought' group are is yet to be ascertained.

It is unfortunate that while there are now a number of independent studies on brain imaging during sexual arousal the resultant descriptions of the areas claimed to be the activated or deactivated are not in agreement. Part of the problem is that different stimuli, duration of stimuli, data handling, and processing protocols have been used but even these do not explain all the differences. Thus, at present, it is not possible to give a detailed and reliable account of brain activation during sexual arousal.

Summary

Normal human sexual function can be characterized simply by its biological mechanisms which are of obvious importance, not least to reproduction.[41] The mechanisms have changed little over the centuries, but their expression as behaviour, when moulded by historical time, social class, ethnic grouping, religion, and society, creates the changing complex concept of human sexuality. Indeed, it has been said that human sexuality is more about fertilizing relationships than eggs! While we have increased hugely our knowledge about many of the mechanisms involved in human sexuality, the impact of a highly successful oral therapy for erectile dysfunction being an obvious example, those of the brain and spinal cord are practically unexplored. What creates human sexual desire and sexual excitement and what causes them to fade away, where in the brain is the pleasure of orgasms created, why do men have a PERT but not women, are just a few of the fascinating questions that remain to be answered.

Further information

Goldstein, I., Meston, C.M., Davis, S.R., *et al.* (eds.) (2006). *Women's sexual function and dysfunction: study, diagnosis and treatment*. Taylor & Francis, London.

Janssen, E. (ed.) (2007). *The psychophysiology of sex*. Indiana University Press, Bloomington, IN.

Janssen, E., Prause, N., and Geer, J. (2007) The sexual response. In *Handbook of Psychophysiology* (3rd edn) (eds. J.T. Cacioppo, G. Tassinary, and G.G. Bernison), pp. 245–66, Cambridge University Press, New York.

Komisaruk, B.R., Beyer-Flores, C., and Whipple, B. (2006). *The science of orgasm*. The Johns Hopkins University Press, Baltimore, MD.

Lue, T.F., Basson, R., Rosen, R., *et al.* (eds.) (2004). *Sexual medicine- sexual dysfunctions in men and women*. Health Publications, Editions 21, Paris, France.

References

1. Morris, D. (1967). *The naked ape*. Jonathan Cape, London.
2. Heltne, P.G. and Marquardt, L.A. (eds.) (1989). *Understanding chimpanzees*. Harvard University Press, Cambridge, MA.
3. Laqueur, T. (1990). *Making sex: body and gender from the Greeks to Freud*. Harvard University Press, Cambridge, MA.
4. Foucault, M. (1980). *The history of sexuality*. Vintage, New York.
5. Tiefer, L. (1992). Historical, scientific, clinical and feminist criticism of the 'the human sexual response cycle' model. *Annual Review of Sex Research*, **2**, 1–23.
6. Gagnon, J.H. and Simon, W. (1973). *Sexual conduct: the social sources of human sexuality*. Aldine Chicago, IL.
7. Money, J. (1986). *Lovemaps: clinical concepts of sexual/erotic health and pathology, paraphilia, and gender transposition in childhood, adolescence, and maturity*. Irvington, New York.
8. Kinsey, A.C., Pomeroy, W.B., and Martin, C.E. (1948). *Sexual behaviour in the human male*. W.B. Saunders, Philadelphia, PA.
9. Kinsey, A.C., Pomeroy, W.B., Martin, C.E., *et al.* (1953). *Sexual behaviour in the human female*. W.B. Saunders, Philadelphia, PA.
10. Clement, U. (1990). Surveys of heterosexual behaviour. *Annual Review of Sex Research*, **1**, 45–74.
11. Einon, D. (1994). Are men more promiscuous than women? *Ethology and Sociobiology*, **15**, 131–43.
12. Berk, R., Abramson, P.R., and Okami, P. (1995). Sexual activities as told in surveys. In *Sexual nature, sexual culture* (eds. P.R. Abramson and S.D. Pinkerton), pp. 371–86. University of Chicago Press, Chicago, IL.
13. Michael, R.T., Gagnon, J.H., Laumann, E.O., *et al.* (1994). *Sex in America—a definitive survey*. Little, Brown, London.
14. Wellings, K., Field, J., Johnson, A.M., *et al.* (1994). *Sexual behaviour in Britain*. Penguin, Harmondsworth.
15. Masters, W.H. and Johnson, V.E. (1966). *Human sexual response*. Little, Brown, Boston, MA.
16. Robinson, P. (1976). *The modernization of sex*. Cornell University Press, Ithaca, NY.
17. Kaplan, H. (1979). *Disorders of sexual desire*. Simon and Schuster, New York.
18. Garde, K. and Lunde, I. (1980). Female sexual behaviour. A study in a random sample of 40-year-old women. *Maturitas*, **2**, 225–40.
19. Basson, R. (2000). *The female sexual response model revisited. Journal of the Society for Obstetrics and Gynecology of Canada*, **22**, 383–7.
20. Buss, D.M. (1994). *The evolution of desire: strategies of human mating*. Basic Books, New York.
21. Trivers, R.L. (1972). Parental investment and sexual selection. In *Sexual selection and the descent of man 1871–1971* (ed. B. Campbell), pp. 136–79. Aldine, Chicago, IL.
22. Grammar, K. (1989). Human courtship behaviour: biological basis and cognitive processing. In *The sociobiology of sexual and reproductive strategies* (eds. A.E. Rasa, C. Vogel, and E. Voland), pp. 147–69. Chapman & Hall, London.
23. Phillips, A. (1993). *On kissing, tickling and being bored*. Harvard University Press, Cambridge, MA.
24. Nicholson, B. (1984). Does kissing aid human bonding by semio-chemical addiction? *The British Journal* of Dermatology, **111**, 623–7.
25. Levin, R.J. (1992). The mechanisms of human female sexual arousal. *Annual Review of Sex Research*, **3**, 1–48.
26. Levin, R.J. (2007). Sexual activity, health and well being—the beneficial roles of coitus and masturbation. *Sexual and Relationship Therapy*, **22**, 135–48.
27. Levin, R.J. (1994). Human male sexuality: appetite and arousal, desire and drive. In *Appetite: neural and behavioural bases* (eds. C.R. Legg and D. Booth), pp. 127–63. Oxford University Press, Oxford.
28. Bremer, J. (1959). *Asexualisation: a follow up of 244 cases*. Macmillan, New York.
29. Wilson, J. (1978). Sexual differentiation. *Annual Review of Physiology*, **40**, 279–306.
30. Meuwissen, I. and Over, R. (1992). Sexual arousal across phases of the human menstrual cycle. *Archives of Sexual Behaviour*, **21**, 165–73.
31. Hutchinson, K.A. (1995). Androgens and sexuality. *The American Journal of Medicine*, **98**(Suppl. 1A), 111S–15S.
32. Gerstenberg, T.C., Levin, R.J., and Wagner, G. (1990). Erection and ejaculation in man. Assessment of the electromyographic activity of the bulbocavernosus and ischiocavernosus muscle. *British Journal of Urology*, **65**, 395–402.
33. Levin, R.J. (2003). *Is prolactin the biological "off switch" for human sexual arousal? Sexual and* Relationship Therapy, **18**, 239–43.
34. Robbins, M.A. and Jensen, G.D. (1978). Multiple orgasm in males. *Journal of Sex Research*, **14**, 21–6.
35. Levin, R.J. (2005). The mechanisms of human ejaculation—a critical analysis. *Sex & Relationship Therapy*, **20**, 123–31.
36. Langfeldt, T. (1990). Early childhood and juvenile sexuality, development and problems. In *Handbook of sexology*, Vol. 7 (ed. M.E. Perry), pp. 179–200. Elsevier, Amsterdam.
37. Levin, R.J. (2004). An orgasm is . . . who defines what an orgasm is? *Sexual & Relationship Therapy*, **19**, 101–7.
38. O'Connell, H.E., Hutson, J.M., Anderson, C.R., *et al.* (1998). Anatomical relation between urethra and clitoris. *The Journal of Urology*, **159**, 1892–7.
39. van Turnhout, A.A.W.M., Hage, J.J., and van Diest, P.J. (1995). The female corpus spongiosum revisited. *Acta Obstetrica Scandinavica*, **74**, 762–71.
40. Levin, R.J. (1998). Sex and the human female reproductive tract-what really happens during and after coitus. *International Journal of Impotence Research*, **10**(Suppl. 1), S14–21.
41. Levin, R.J. (2005). Sexual arousal—its physiological roles in human reproduction. *Annual Review of Sex Research*, **16**, 154–89.
42. Riley, A.J., Lees, W.R., and Riley, E.J. (1992). An ultrasound study of human coitus. In *Sex matters* (eds. W. Bezemer, P. Cohen-Kettenis, and K. Slob), pp. 29–32. Excerpta Medica, Amsterdam.
43. Grafenberg, E. (1960). The role of urethra in female orgasm. *International Journal of Sexology*, **3**, 145–8.
44. Ladas, A.K., Whipple, B., and Perry, J.D. (2005). *The G spot and other recent discoveries about human sexuality*. Henry Holt & Company Ltd., New York.
45. Minh, H.-N., Smadja, A., and De Sigalony, J.P. (1979). Le fascia de Halban: son role dans la physiologie sexuelle. *Gynaecologie*, **30**, 267–73.
46. Hoch, Z. (1986). Vaginal erotic sensitivity by sexological examination. *Acta Obstetrica et Gynaecologica Scandinavica*, **65**, 767–73.
47. Holstege, G., Geogiadis, J.R., Paans, A.M.J., *et al.* (2005). Brain activation during human male ejaculation. *Journal of Neuroscience*, **23**, 9185–93.
48. Komisaruk, B. and Whipple, B. (2005). Functional MRI of the brain during orgasm in women. *Annual Review of Sex Research*, **16**, 62–86.
49. Ferretti, A., Caulo, M., Del Gratta, C., *et al.* (2005). Dynamics of male sexual arousal: distinct components of brain activiation revealed by fMRI. *NeuroImage*, **26**, 1086–96.

4.11.2 The sexual dysfunctions

Cynthia A. Graham and John Bancroft

Introduction

Sexual relationships are central to the lives of most of us. The sexual component of those relationships can go wrong in various ways. This may be secondary to other difficulties in the relationship, mental health problems, specific sexual vulnerabilities of the individual, or the impact of disease or medication on sexual response. This chapter will describe the more common sexual problems and their prevalence. Evidence related to aetiology of sexual problems and treatment evaluation will be briefly reviewed. In the final section of the chapter, guidelines for the assessment and practical management of sexual problems will be presented.

Historical aspects and some basic concepts

Since 1970, when Masters and Johnson[1] published their groundbreaking book on the treatment of 'human sexual inadequacy', there have been two lines of development in this field, relatively detached from each other until recently: psychological methods of treatment, collectively known as 'sex therapy' and medical interventions, initially focused on erectile problems in men.

The involvement of the medical profession has been substantial, although predominantly involving urologists. Initially there were surgical procedures to implant penile splints or to improve the vascular supply to the penis, and the use of vacuum devices to induce erection mechanically. This was followed by the discovery that injection of smooth muscle relaxants, such as papaverine, phentolamine, or prostaglandin into the erectile tissues of the corpora cavernosa induced erections. Self-injections became widely prescribed. To avoid the need for penile injections, which were not popular among male patients, preparations of prostaglandin for intra-urethral administration became available. This era of medical intervention was characterized by a veritable industry of investigative procedures in attempts to identify local causes for erectile dysfunction (ED). Erectile problems were clearly differentiated into 'organic' and 'psychogenic' subtypes. There was, however, a notable lack of attention to how the brain and psychological processes interacted with these peripheral mechanisms.

Then came the 'Viagra revolution' in the early 1990s. The first oral phosphodiesterase 5 (PDE-5) inhibitor, sildenafil (Viagra®), was found to be effective in enhancing erectile response to sexual stimulation when taken about 1 h before sexual activity. This led to the next phase in the 'medicalization' of male sexual dysfunction, with a shift to the primary care physician as the principle source of treatment and a dramatic reduction in the amount of diagnostic assessment.

The progress of sex therapy has been limited since Masters and Johnson.[1] It has continued to be used, with various adaptations of the original 'sensate focus' approach, incorporating principles of psychoanalytic techniques[2] and cognitive behaviour therapy.[3] The main shortcoming has been inadequate outcome research on the efficacy of these methods.

The next phase in this recent history followed the commercial success of PDE-5 inhibitors for men, with an inevitable quest for a 'Viagra for women'. This has so far proved elusive, but has confronted the 'sexual medicine' community with the complexity of women's sexuality and the need to conceptualize it differently to the sexuality of men.

At the same time, evidence has emerged that treatment of ED with sildenafil and more recent PDE-5 inhibitors, although initially successful in the majority, was being discontinued by a substantial proportion of men.[4] In addition, the female partners of men taking these drugs do not always welcome the associated changes in the sexual relationships.[5] We are now moving into the most recent phase where the 'psychological' and 'organic' approaches, and the professional groups that have been identified with them, have started to interact. There is increasing recognition of the need to integrate psychological and medical methods of treatment,[6,7] but with the important proviso that, at least initially, treatment should focus on the couple and not the individual.

One important aspect of this evolving story is how we define a 'sexual dysfunction', with connotations of abnormal or impaired function, and how it is distinguished from a 'sexual problem' in a more general sense. This issue was epitomized by a publication in the *Journal of the American Medical Association* on the epidemiology of 'sexual dysfunction'.[8] In this widely cited paper, 43 per cent of women and 31 per cent of men were identified as having a 'sexual dysfunction', described as 'a largely uninvestigated yet significant public health problem' (p. 544). The authors commented, 'With the affected population rarely receiving medical therapy for sexual dysfunction, service delivery efforts should be augmented to target high-risk populations'(p. 544).

This dramatic example of 'medicalization', based on extremely limited information from a national survey not designed to assess sexual dysfunction, was effectively challenged by Mercer and colleagues, using data from the UK National Survey of Sexual Attitudes and Lifestyles (NATSAL).[9] This used exactly the same questions as in the Laumann *et al.* study,[8] with the important difference that it was more specific about duration of problems, asking whether particular problems had lasted 'at least 1 month', or 'at least 6 months' during the last year (Laumann *et al.* had asked if symptoms had occurred 'for several months or more' during the last year). Overall, 53.8 per cent of women and 34.8 per cent of men reported at least one sexual problem lasting at least 1 month during the previous year. In contrast, the prevalence of problems lasting 'at least 6 months in the previous year' was 15.6 per cent for women and 6.2 per cent for men. This showed that transient problems were very common, more persistent ones much less so. In both the American and the British study, such problems were related to other problems in the participants' lives, particularly involving impaired mental health (e.g. depression), relationship problems, or significant life stresses.

Relevant to the question of when a sexual problem becomes a 'dysfunction' is a theoretical approach, called the 'Dual Control Model', developed at the Kinsey Institute. This postulates that sexual response results from an interaction between excitation and inhibition, involving relatively discrete neurophysiological systems in the brain.[10] A central assumption of the model is that individuals vary in their propensity for both sexual excitation and sexual inhibition and that 'normal' levels of inhibition are adaptive, reducing sexual responsiveness in circumstances where sexual

activity is best avoided. It is predicted that high levels of inhibition may be associated with vulnerability to sexual dysfunction and low levels with an increased likelihood of engaging in high-risk sexual behaviour.[10]

This faces us with the seemingly obvious but fundamental challenge of deciding whether a loss of sexual interest or responsiveness is an understandable or even adaptive reaction to current circumstances, or is a result of 'malfunction' of the sexual response system, which can appropriately be called a 'sexual dysfunction'. This challenge is also central to the relatively new phase of integrated treatment, in which assessment identifies the key factors causing the sexual problem and how they should best be treated. A strategy for carrying out such assessment, which we have called the 'three windows approach', will be outlined below.

Clinical features of sexual problems

Sexual problems in men

The most common problems presented by men are ED and premature ejaculation (PE). Delayed or absent ejaculation is a relatively infrequent complaint. Low sexual desire may be the presenting problem, although in most cases this is combined with ED, and it is not always clear which came first.

(a) Erectile problems

Penile erection is a tangible and fundamental component of a man's experience of sexual arousal and the lack of erection in a sexual situation often has significant negative effects. Irrespective of whether or not there are peripheral explanations for impaired erections (e.g. vascular disease), the reactions of the man and his partner have a major influence on how problematic the erectile difficulty becomes. Erectile difficulties vary in severity; in some men the problem only occurs on a proportion of occasions of sexual activity. The difficulty may be in getting an erection or in maintaining it long enough for satisfactory sexual intercourse.

(b) Low sexual desire

For many men, sexual desire is linked with erectile responsiveness. Many men with low sexual desire also report a reduction in 'spontaneous' erections. However, a man can experience low sexual desire without having any erectile difficulties, although he may require more direct tactile stimulation to achieve erections.

(c) Premature ejaculation (PE)

Ejaculation results from a combination of orgasm and seminal emission, with muscular contractions as part of the orgasmic response resulting in expulsion of the seminal emission. PE is essentially a problem when the man is unable to delay orgasm and ejaculation as he would wish. Not surprisingly, this has led to considerable inconsistencies of definition in the literature. In severe cases, emission occurs before vaginal entry and the orgasmic component may be so reduced that the usual muscle spasms do not occur, resulting in semen seeping out of the urethra rather than being 'ejaculated'.

Premature ejaculation has been categorized as 'primary' (i.e. lifelong) or 'secondary'. Secondary PE is often confounded by erectile problems. If a man is taking a long time to get an erection, he may reach the stimulus intensity required for ejaculation before or soon after erection is achieved.

(d) Delayed ejaculation

Delayed or absent ejaculation occurs in men, although it is much less common than rapid ejaculation. A man might have difficulty ejaculating only during sexual activity with his partner and in some cases only during penetrative intercourse, or the problem may be evident even during masturbation. Delayed or absent ejaculation is a common side effect of selective serotonin re-uptake inhibitor (SSRI) medications, which often prevents orgasm in women as well, suggesting that the primary effect of such drugs is on the triggering of orgasm.

(e) Pain during sexual response

Pain may be associated with prolonged sexual arousal not terminated by ejaculation/orgasm. Such pain is usually experienced in the testes. Pain felt in the urethra may occur during ejaculation. Neither problem is common.

Sexual problems in women

(a) Loss of sexual arousal and/or desire

Most surveys have suggested that low sexual desire is the most common sexual problem reported by women. However, low sexual desire is a heterogeneous problem category and the relationship between sexual arousal and sexual desire in women is particularly complex. Many women do not differentiate between 'arousal' and 'desire'[11] and awareness of 'desire' is usually accompanied by some degree of central arousal, whether or not any genital response is perceived.[12] It has been argued that sexual desire in women is much more likely to be 'receptive' and triggered by a desire for intimacy with one's partner.[13] It is therefore not surprising that there is considerable overlap or comorbidity between problems related to sexual arousal and desire in women.[8]

Although traditionally seen as the counterpart to penile erection in men, vaginal response is not central to the experience of sexual arousal in women. Vaginal dryness may be a problem because of the likelihood of discomfort or pain with intercourse when the vagina is not adequately lubricated, but this symptom does not necessarily indicate lack of arousal. Conversely, a woman may experience lack of sexual arousal and yet have vaginal lubrication. The relevance of vaginal response to sexual arousal in women therefore remains unclear. An increase in vaginal blood flow has been consistently demonstrated in women reacting to sexual stimuli, whether or not they find the sexual stimulus appealing; this led Laan and Everaerd[12] to call this an 'automatic' response. There is no obvious counterpart to this in men. Tumescence of the clitoris, on the other hand, may be more directly comparable to male genital response but this is less easily assessed and less clearly perceived by women, compared with penile erection in men.

Persistent genital arousal disorder (PGAD) is a recently recognized but fairly uncommon sexual problem in women. It is characterized by genital and breast vasocongestion and sensitivity which persists for hours or days and is only temporarily relieved by orgasm; genital sensations are unaccompanied by any subjective sense of sexual desire and excitement but instead are perceived as intrusive. There is no male equivalent to this problem, probably because the post-orgasmic refractory period is more substantial in men.

Much less frequent than loss of sexual arousal or desire is extreme aversion to, and avoidance of all sexual contact with, a sexual partner. This can occur in women and in men.

(b) Problems with orgasm

Difficulty in achieving orgasm is not uncommon in women. Often this is situational in that orgasm is possible with masturbation, but not during sexual interaction with the partner. The capacity to experience orgasm varies considerably across women. Some women reach orgasm easily if sufficient arousal occurs, others may require more specific or more intense stimulation, and an estimated 10–15 per cent are unable to experience orgasm throughout their lives.[14] In identifying a problem as primarily orgasmic, one needs to first establish that appropriate sexual arousal has occurred.

(c) Problems with sexual pain and vaginismus

Pain during attempted or complete vaginal entry (dyspareunia) is a common sexual problem in women with a wide range of possible causes. Sexual pain is also frequently associated with lack of sexual desire and/or arousal. Vaginismus has traditionally been defined as recurrent or persistent involuntary spasm of the musculature of the outer third of the vagina that makes vaginal penetration difficult or impossible. This definition has recently been questioned because of a lack of empirical evidence that vaginal spasms occur in women diagnosed with vaginismus.[15] Vulvar vestibulitis syndrome (VVS) is a condition associated with pain on touching the labia or vaginal introitus. The question of whether these are primarily 'sexual' or pain disorders is currently under debate.[16]

Classification of sexual problems

DSM-IV and ICD classification

The current Diagnostic and Statistical Manual of Mental Disorders (DSM) classification[17] defines sexual dysfunction as characterized by 'disturbance in sexual desire and in the psychophysiological changes that characterize the sexual response cycle and cause marked distress and interpersonal difficulty' (p. 493). There is Hypoactive Sexual Desire Disorder and Sexual Aversion Disorder, defined in the same way for men and women. Female Sexual Arousal Disorder is defined as 'a persistent or recurrent inability to attain, or to maintain until completion of the sexual activity, an adequate lubrication-swelling response of sexual excitement' (p. 502) and, for the male version, erection is the relevant response. Orgasmic Disorder (i.e. delayed or absent orgasm) and Dyspareunia are defined in basically the same way for men and women. Vaginismus is a specifically female diagnosis and Premature Ejaculation an exclusively male disorder.

The International Statistical Classification of Diseases and Related Health Problems[18] covers sexual dysfunctions in one and a half pages, compared with nearly 30 pages in the DSM. The basic categories of dysfunction are similar to those of DSM, but there are few, if any, actual diagnostic criteria given for any of the dysfunctions. ICD-10 also does not require that personal distress or interpersonal problems are present for a diagnosis to be made. Instead, there is the statement 'sexual dysfunction covers the various ways in which an individual is unable to participate in a sexual relationship as he or she would wish' (p. 355).

There has been increasing dissatisfaction with the current classification of sexual dysfunction for women.[19,20]. Major areas of criticism include the high comorbidity between diagnoses of sexual dysfunction and the 'genital' focus of the diagnostic criteria and concomitant neglect of psychological and relationship factors. In response, alternative models of sexual response[13] and women-centred definitions of sexual problems[21] have been proposed.

Epidemiology

There have been at least 12 representative community-based surveys that have assessed the prevalence of sexual problems.[22] Prevalence rates for specific problems vary considerably and whereas several of the studies claim to be reporting prevalence of sexual 'dysfunctions',[8] it is now accepted that the detailed clinical assessment required to identify a sexual dysfunction cannot be undertaken by large-scale surveys.[23] Consequently, more recent surveys have used terms such as 'problems'[9,24] or 'difficulties'.[25] Variability in reported prevalence rates can in part be attributed to variations in how sexual problems are defined, their duration, and how and whether 'distress' about changes in sexual functioning is assessed. In studies of female sexual problems, there has been only limited overlap between what women perceive as problematic and what the researcher or clinician identifies as a problem.[26,27] The variability of prevalence rates is shown in Tables 4.11.2.1 and 4.11.2.2, which compare a number of population-based surveys involving women and men.

There has been more consistency across studies in the associations found between factors of possible aetiological relevance and sexual functioning. In women, sexual problems are more frequent in those with mental health problems and relationship difficulties.[8] In a survey of women aged 20–65 years, all in heterosexual relationships, 24.4 per cent reported marked distress about their sexual relationship and/or their own sexuality.[26] The best predictors of distress were markers of mental health and the quality of the emotional relationship with the partner. Physical aspects of sexual response in women such as arousal and orgasm were poor predictors of distress.

In men, age has a predictable negative effect on erectile function. In one study, complete erectile failure was reported by 5 per cent of men at age 40 and 25 per cent at age 70.[28] A similar age effect is found with loss of sexual desire. In another study, absence of any sexual desire was reported by 2 per cent of men aged 45–59 and 18.2 per cent aged 75+.[29] Contrary to popular belief, PE does not show a clear negative relationship with age.

The association between age and sexual problems in women is more complex. Whereas level of sexual interest typically decreases with age, older women are less likely to regard this as a problem.[26] Older women are much less likely to be in a sexual relationship than older men; the presence or absence of a partner also seems to influence women's sexuality to a greater extent than it does for men.[29] The impact of the menopause is also complex. Although the post-menopausal decline in oestrogens is relevant to vaginal lubrication, other factors such as mental health and the quality of the sexual relationship are more important determinants of sexual well-being.[30]

Aetiology

Before considering the factors that can cause sexual problems, it is worth underlining the important way that sexual function differs from most other physiological response systems. Although involving physiological mechanisms, sexual responses are most often experienced in the context of a relationship. This highlights the

Table 4.11.2.1 Prevalence of specific female problems (%) found in seven community-based surveys

Study	Low sexual interest	Impaired arousal	Impaired lubrication	Impaired orgasm	Pain	Total (one or more problems)
Richters *et al.* (2003)[25]1						
At least 1 month	54.8	–	23.9	28.6	20.3	
Mercer *et al.* (2003)[9]1						
At least 1 month	40.6	–	9.2	14.4	11.8	53.8
At least 6 months	10.2	–	2.6	3.7	3.4	15.6
Bancroft *et al.* (2003)[26]3	7.2	12.2	31.2	9.3	3.3	45
Laumann *et al.* (1999)[8]1	31.6	–	20.6	25.7	15.5	43[a]
Fugl-Meyer and Fugl-Meyer (1999)[59]1	33.0	–	12.0	22.0	6.0	47
Dunn *et al.* (1999)[24]2	–	17.0	28.0	27.0	18.0	41
Osborne *et al.* (1988)[37]2	17.0	–	17.0	16.0	8.0	33

[a] Includes additional problem categories not shown in this table.

[1] During last year.

[2] During last 3 months.

[3] During last month.

importance of keeping the interactive relationship components in mind when trying to assess and treat sexual problems. Socio-cultural factors are also crucial to understanding how sexual problems are experienced. Much of the focus in medical treatments of sexual problems has been on the individual patient, with relationship and socio-cultural aspects largely ignored. The more specific aetiological factors can now be considered using the 'three windows approach'.[26]

The first window—the current situation

Through the first window, a variety of factors in the individual's current relationship and situation may be relevant. Relationship problems, particularly resentment and insecurity within a relationship, are of particular importance. For many individuals, feeling secure and being able to 'let go' are necessary for them to enjoy sex. Other factors that may be important include: poor communication between partners about their sexual feelings and needs; misunderstandings and lack of information; unsuitable circumstances and lack of time e.g. fatigue, lack of privacy, and work pressures; concerns about pregnancy or about sexually transmitted diseases; and low self-esteem and poor body image.

Various mechanisms may mediate the effects of the above situational factors on sexual functioning. Reactive inhibition, as postulated by the Dual Control Model,[10] may well be involved in those circumstances associated with a negative emotional response. With stress and fatigue, the mechanisms are less clear, but may entail a transient reduction in the capacity for excitation (i.e. sexual arousability).

Table 4.11.2.2 Prevalence of specific male problems (%) found in five community-based surveys

Study	Low sexual interest	Erectile problem	Premature ejaculation	Delayed ejaculation	Total (one or more problems)
Richters *et al.* (2003)[25]1					
At least 1 month	24.9	9.5	23.8	6.3	–
Mercer *et al.* (2003)[9]1					
At least 1 month	17.1	5.8	11.7	5.3	34.8
At least 6 months	1.8	0.8	2.9	0.7	6.2
Laumann *et al.* (1999)[8]1	14.6	10.2	30.6	7.8	31.0[a]
Fugl-Meyer and Fugl- Meyer (1999)[59]1	16.0	5.0	9.0	–	–
Dunn *et al.* (1999)[24]2	–	26.0	14.0	–	–

[a] Includes additional problem categories not shown in this table.

[1] During last year.

[2] During last 3 months.

The second window—vulnerability of the individual

Although a wide range of factors can impact on our sexuality, it is also clear that individuals vary substantially in the extent to which they are affected by such factors, particularly in terms of an associated inhibition of sexual response. Such vulnerabilities are likely to have been evident in earlier episodes in the current relationship or in earlier relationships.

(a) Negative attitudes

Long-standing attitudes, usually stemming from childhood, that sex is inherently 'bad' or immoral are likely to interfere with an individual's ability to become involved in and enjoy a sexual relationship.

(b) Need to maintain self-control

In some individuals, a difficulty in 'letting go' sexually reflects a more general need to maintain self-control, particularly in the presence of another person.

(c) Earlier experience of sexual abuse or trauma

There is now an extensive literature on the impact of sexual abuse on subsequent sexual adjustment. Whereas the mediating mechanisms are not well understood and are likely to be complex and varied, a history of such experience should be regarded as potentially relevant to current sexual difficulties.

(d) Propensity to sexual inhibition

The Dual Control Model, discussed earlier, has led to psychometrically validated measures of propensity to sexual excitation and inhibition in men[31] and women.[32] These are new measures and research exploring their relevance to vulnerability to sexual problems has only recently started. However, using these measures, close to normal distributions of scores for both sexual excitation and inhibition proneness have been reported in men[31] and women.[32] As predicted, a clear association between low sexual excitation and/or high sexual inhibition propensity and erectile problems in men has been found, but no association with PE.[33] One study explored a possible genetic basis for such individual variability in men and found evidence of heritability for propensity to sexual inhibition but not excitation.[34] In women, a strong relationship between sexual inhibition scores and reports of sexual problems was found.[35] Particularly important was one inhibition subscale (labelled 'arousal contingency') that reflects susceptibility for sexual arousal to be easily affected by situational factors e.g. if the circumstances are not 'just right' or if the woman is distracted. Although further research is needed, these measures of sexual inhibition and excitation may prove valuable in explaining patterns of impaired sexual response and helping in the selection of appropriate treatment. As yet no other personality-related or trait measure has been shown to have clinical value in this respect.

The third window—health-related factors that alter sexual function

The variety of health-related factors are considered under three headings: mental health, physical health, and sexual side effects of medication.

(a) Mental health and sexuality

Psychiatric problems are commonly associated with sexual problems.[36,37] Reduction in sexual interest, and to some extent sexual arousability, is generally accepted as a common symptom of depressive illness.[38] In contrast, sexual interest tends to be increased in states of elevated mood such as hypomania.[39] In a study of men and women with primary loss of sexual interest, the large majority had experienced previous affective disorder, with reduced sexual desire being established during, and persisting after, one of these earlier depressive episodes.[40]

We would expect reactive inhibition of sexual interest or response in circumstances eliciting negative mood (i.e. 'reactive depression'). With more endogenous depression, however, reduction of the excitatory component of sexual response is also evident. This is shown clearly in the impairment of nocturnal penile tumescence (NPT) that typically occurs with depressive illness. NPT, while not completely understood, probably results from a 'switching off' of inhibitory tone during REM sleep. This therefore allows expression of 'excitatory tone', presumably impaired in depression with associated metabolic changes.

With anxiety, the clinical evidence is much more limited. Higher rates of sexual dysfunction in patients with anxiety disorders have been reported.[41] In the Zurich Cohort Study, a longitudinal study of men and women aged 20–35 years, loss of sexual interest was more prevalent in patients with generalized anxiety disorder, but was not associated with panic disorder, agoraphobia, or social phobia.[42] In another study, patients with panic disorder were more likely to report sexual problems, particularly sexual aversion, than social phobics; PE was the most common sexual problem in men with social phobias.[43]

Non-clinical, community-based studies have also demonstrated a relationship between mood and sexuality. In the Massachusetts Male Aging Study, an association was reported between ED and depressive symptoms, after controlling for other potential confounding variables such as age and physical health.[44] Angst[42] found a relationship between depression and loss of sexual interest, particularly marked in women. In a US survey of heterosexual women, scores on a brief measure of mental health were strongly predictive of women's distress about their sexual relationship and their own sexuality.[26]

Although most studies have focused on negative effects of mood disorders on sexual interest and response, there is evidence that a minority of individuals experience increased sexual interest during negative mood states. Angst[42] reported that among depressed men, 25.7 per cent reported decreased, and 23.3 per cent increased, sexual interest, compared with 11.1 per cent and 6.9 per cent, respectively, of their non-depressed group. In contrast, among depressed women, 35.3 per cent reported decreased sexual interest and only 8.8 per cent reported increased interest (compared with 31.6 per cent and 1.7 per cent, respectively, of their non-depressed group). This paradoxical pattern of increased sexual interest during negative mood states has also been reported in non-clinical samples of men[45] and women.[46] Although the origins of this pattern are not yet understood, it may well be problematic in various ways; for example, associated with sexual risk-taking or leading to sex being used as a mood regulator.

The impact of schizophrenia on sexuality is complex. Given the importance of dopaminergic neurotransmission to various aspects of sexuality, and the fact that most anti-psychotics are dopamine antagonists, one might expect amplification of sexual interest or response in some form, at least during the more florid type one stage of the illness. Sexual thoughts and behaviours are common in

schizophrenia and there may be a relative increase in sexual activity. However, this is typically autoerotic or 'compulsive' without any real 'object-relational' quality.[47] In view of the effects of this illness on interpersonal functioning, this is not surprising. Early studies suggested that loss of sexual interest was less likely in schizophrenia than in other types of psychiatric illness, although female schizophrenics were more likely to report a reduction in sexual interest than males.[47] 'Pre-schizophrenic personality', evident in the history of many cases, may also be associated with an interference with normal sexual development.

(b) Physical health and sexuality

The impact of poor physical health on sexuality may be relatively non-specific. For example, loss of well-being and energy associated with chronic illness is likely to cause reduced sexual interest and arousability. Psychological reactions to the illness or condition (e.g. the effects of breast cancer on a woman's body image and hence her sexual enjoyment) may also be important. In addition there are a variety of ways in which the health problem can directly affect sexual interest and/or response. In most cases we find the evidence much clearer in men than in women.

(i) Damage to the neural control of genital response

This can involve peripheral mechanisms (e.g. autonomic neuropathy and peripheral neuropathy) or disease in the spinal cord (e.g. multiple sclerosis). Injury or surgery causing nerve damage may be involved (e.g. spinal cord injury, prostatectomy, or hysterectomy). The most likely consequences are ED in men and impaired vaginal response in women. There is no clear relationship between neural damage and PE but interference with normal seminal emission can occur, with resulting loss of ejaculation or retrograde ejaculation when, due to nerve damage to the pelvic muscles, the seminal fluid passes backwards into the bladder during ejaculation. Brain abnormalities, such as epilepsy or cerebral tumour, can affect central control of sexual response, the precise effect depending on the site of the abnormality or tumour. In some cases the result is loss of sexual interest and arousability; rarely there is disinhibition of sexual behaviour.

(ii) Impairment of vascular supply of the genitalia

Genital response is dependent on increased arterial inflow as well as alteration in venous outflow. Vascular impairment can result from peripheral vascular disease, affecting the small vessels, or obstruction in a main artery. ED is also common in men who have ischaemic heart disease.

(iii) Alteration of endocrine mechanisms affecting sexual interest, arousal and response

In males any cause of lowered testosterone (T) levels is likely to produce loss of sexual interest and, to a varying extent, impairment of erectile response. Hypogonadism, if severe enough, will also be associated with loss of seminal emission (and usually orgasm).

In women, lack of oestrogen is associated with impaired vaginal lubrication. The effects of sex steroids, either oestrogens or androgens, on sexual interest and arousability, are much less predictable. The evidence is consistent with there being a proportion of women who depend on T for their normal level of sexual interest, but there are many women who can experience substantial reduction in T without obvious adverse sexual effects. The role of oestrogen in the more central aspects of women's sexuality also remains unclear.[48]

Hyperprolactinaemia, usually resulting from pituitary adenomas and hypersecretion of prolactin, may be associated with loss of sexual interest, and in men, with ED. However, this is not always the case. The precise role of prolactin in human sexuality is not understood. Its central control by dopamine and serotonin contributes to the complexity. Adverse effects of hyperprolactinaemia on sexuality are probably more likely in men than women.

(iv) Metabolic disorders

Various forms of metabolic disturbance, such as that associated with hepatic or renal disease, may be associated with adverse effects on sexual interest and response, though the mediating mechanisms are not understood.

Some diseases affect sexuality through more than one of the above mechanisms. Diabetes mellitus is a good example. ED has long been recognized as a complication of diabetes and is more common in Type 1 than Type 2 diabetes. In diabetic men, ED can result from small vessel vascular disease and also autonomic and peripheral neuropathology. Diabetes is associated with hypogonadism in men. Lowered sexual interest and impairment of genital response have also been reported in diabetic women.

(c) Side effects of medication

In considering the complexities of pharmacological effects on sexual function, it is helpful to distinguish between excitatory and inhibitory mechanisms. In the brain, sexual excitation depends to a considerable extent on dopamine (DA) and noradrenaline (NA). DA is involved in the 'incentive motivation system' and is hence relevant to sexual interest. It is the D2 receptor that is most relevant to these sexual effects. It is also involved in the hypothalamic control of genital response. NA, when acting centrally, is involved in central arousal, a key component of sexual arousal though not specific to it. In the periphery, depending on the receptor involved, NA can have inhibitory or enhancing effects on smooth muscle relaxation, a critical aspect of genital response. There are three principal NA receptors of relevance: (i) alpha-1, which is post-synaptic and mediates smooth muscle contractions, inhibiting genital response, (ii) alpha-2 which is principally pre-synaptic, where it increases re-uptake of NA and hence reduces the amount of NA at the synapse, and (iii) beta-2, a peripheral receptor which mediates a relaxing effect of NA on smooth muscle in the urogenital system and elsewhere. Serotonin (5-HT) has a central role in the inhibitory system. The two most relevant receptors are the 5-HT2 receptor, which mediates the inhibitory effects and the 5-HT1A receptor, which is pre-synaptic and, comparable to the alpha-2 receptor, reduces 5-HT transmission. We can now consider the main adverse sexual effects resulting from medication; for a review, see.[49]

(d) Anti-depressants

The clearest examples are the SSRIs which, by inhibiting the 5HT-1A receptor, increase serotonergic transmission. The most predictable effect, in both women and men, is inhibition of orgasm and ejaculation. This effect is being exploited in the treatment of PE, with the development of short-acting SSRIs which can be used as needed prior to sexual activity. Other negative effects include reduced sexual interest and arousability, though they are less predictable and not always easy to distinguish from the sexual effects of the affective disorder being treated. Tricylic anti-depressants also commonly produce sexual side effects. Two anti-depressants which

have less sexual side effects are bupropion, which is metabolized into a DA and NA re-uptake inhibitor, and nefadazone, which has a 5-HT2 antagonist effect.

(e) Anti-psychotic medication

These drugs, used for the treatment of schizophrenia and other psychotic disorders, involve a balance of DA antagonist and 5-HT agonist effects. Sexual side effects occur in around 60 per cent of men and 30–90 per cent of women using these drugs. The most common side effects are ED and ejaculatory difficulties in men and orgasmic dysfunction in women. It is not clear to what extent these effects are due to the DA antagonist or 5-HT agonist effects, or a balance of the two.

(f) Anti-hypertensive medication

Many drugs used to treat hypertension interfere with male sexual response. In the case of alpha-1 antagonists (e.g. prazosin), the peripheral effects would be expected to enhance erection, by reducing adrenergic contraction of erectile smooth muscle and indeed priapism is an occasional problem. There is a low incidence of sexual side effects, mainly failure of ejaculation, and central alpha-1 blockade might reduce central arousal. Beta-blockers (e.g. propanalol), by blocking smooth muscle relaxation and leaving the alpha-1 induced contraction unopposed, commonly cause ED. Centrally acting anti-hypertensives (e.g. guanethidine) also interfere with sexual response, impairing erection, and blocking ejaculation. Clonidine, an alpha-2 agonist, causes erectile problems in about 25 per cent of cases. In this case the principal effect is likely to be reduced central arousal. For a review of the effects of anti-hypertensive medication on male sexual response, see.[50]

The effects of drugs on women's sexuality have received far less attention. Research has mostly focused on difficulties in achieving orgasm. As orgasm only occurs after sufficient sexual arousal, these effects may reflect impairment of arousal but this has not been adequately assessed. Steroidal contraceptives, although associated with markedly reduced levels of free T, decrease sexual interest only in a minority of women.[51] This, again, has been little studied and it remains possible that the adverse effects on sexual interest only occur in those women whose sexuality is 'T dependent'.[51]

Management of sexual problems

In this section we will briefly describe the principles of sex therapy and the main forms of pharmacological treatment, followed by an outline of the process of assessment and selection of a suitable treatment plan. The growing awareness of the limitations of pharmacological treatments administered on an individual basis, discussed earlier, has led to recognition that an integration of psychological and pharmacological methods, with emphasis on the couple, may be the most appropriate treatment model in most cases.

Sex therapy

Although the approach first introduced by Masters and Johnson[1] has been adapted in various ways, the core treatment techniques remain. Originally developed for helping couples, the techniques can be modified for use with individuals and with same-sex as well as heterosexual couples.

The key elements of the therapeutic process are:

(a) clearly defined tasks are given and the couple asked to attempt them before the next therapy session

(b) those attempts, and any difficulties encountered, are examined in detail

(c) attitudes, feelings, and conflicts that make the tasks difficult to carry out are identified

(d) these are modified or resolved so that subsequent achievement of the tasks becomes possible

(e) the next tasks are set, and so on

The tasks are mostly behavioural in nature. They are chosen to facilitate the identification of relevant issues but in some cases are sufficient in themselves to produce change. The behavioural programme is in three parts. In the first part the couple are asked to avoid any direct genital touching or stimulation and to focus on non-genital contact, alternating who initiates and who does the touching. These first, non-genital steps are effective in identifying important issues in the relationship, such as lack of trust or counter-productive stereotypical attitudes (e.g. once a man is aroused, he can't be expected not to have intercourse). Once this stage can be carried out satisfactorily, and related problematic issues dealt with, the programme moves on to the second part, which allows genital touching to be combined with non-genital touching, with penile–vaginal intercourse still 'out of bounds'. In this second part more intra-personal problems, such as long-standing negative attitudes about sex, or the sequelae of earlier sexual trauma are likely to emerge. In the third part, a gradual approach to vaginal–penile contact and insertion is undertaken. Here the most relevant issues are 'performance anxiety' and fear of pain.

As the behavioural tasks reveal key issues that need to be resolved before moving on to the next stage, a variety of psychotherapeutic approaches, including cognitive behavioural techniques, can be utilized. Although the stage at which particular issues emerge does vary from case to case, there is a tendency for problems identified through the 'first window', particularly those related to lack of trust and unresolved resentment, to appear during the first stage of the programme. Intra-personal issues (e.g. as seen through the 'second window') are more likely to be recognized during the second stage.

The goals of therapy include helping the individual to accept and feel comfortable with his or her sexuality and helping the couple to establish trust and emotional security and to enhance the enjoyment and intimacy of their sexual interaction. An important point is that these goals do not include reversal of specific sexual dysfunctions. There are exceptions; for example, there are specific behavioural techniques to deal with PE and vaginismus (for details of these techniques, see Bancroft.[52] However, the overriding principle is that, assuming there is no abnormality of the basic physiological mechanisms involved in sexual response, normal sexual function (in terms of sexual desire, arousal, and genital response) will return once the above goals are achieved. In cases where impairment of physiological mechanisms does exist, the above goals of sex therapy are still helpful and integrate well with the use of pharmacological treatment.

Practical aspects

Although sex therapy varies in duration, 12 sessions over 4 to 5 months is typical. The therapist adjusts to the particular needs of the individual or couple. Treatment begins weekly with the interval between sessions extended once major issues like unexpressed

resentment or communication problems have been dealt with. The last two or three sessions are spaced out over a few months so that the couple have an opportunity to consolidate their progress and cope with any setbacks before termination. Open-ended arrangements about length of treatment are best avoided. A specified number of sessions are agreed on at the outset with the proviso that progress will be assessed and a decision made on that basis whether to continue for longer.

A more complete description of the sex therapy process is provided by Bancroft.[52]

Pharmacological treatments for men

(a) Erectile dysfunction. Phosphodiesterase 5 inhibitors

The most important development in this field was the serendipitous discovery that sildenafil (Viagra®), a PDE-5 inhibitor, enhances erectile response. Dose titration is usually required, with available tablets containing 25, 50, or 100 mg. For more severe cases of ED, 100 mg taken about 1 hour before sexual activity is required. The maximal effects last for about 4 h. The most common side effects, that are dose-related, are headache, flushing, and dyspepsia.

Two further PDE-5 inhibitors have been developed and are now in use: tadalafil (Cialis®) and vardenafil (Levitra®). They are both comparable to sildenafil in their efficacy, the main differences being their speed and duration of action. Tadalafil should be taken at least 30 min before sexual activity and has a half-life of 17.5 h; the treatment effect can persist for 24–36 h. Vardenafil is pharmacokinetically similar to sildenafil, but is more potent, with a dose range 5 mg to 20 mg. They are similar in their side effects, though the duration of these relate to the half-life of the drug. For all three drugs, the most important and dangerous drug interaction is with nitrates used for ischaemic heart disease; this is a strong contraindication for the use of PDE-5 inhibitors.

There is now an extensive literature demonstrating the efficacy of sildenafil in the treatment of ED, which has been well reviewed by Rosen and McKenna.[53] Overall, treatment is effective in about 75 per cent of cases. The effectiveness of tadalafil and vardenafil has also been demonstrated.

In spite of their efficacy, the continuing use of PDE-5 inhibitors appears to be limited. In a recent large study done in eight countries, 2912 men identified with ED were followed up; 58 per cent of them had sought medical help for the ED, but only 16 per cent of men maintained their use of PDE-5 inhibitors.[7] Various reasons were given for discontinuation, including lack of appropriate information from the physicians, fear of side effects, partner concerns, and distrust of medications.

(b) Anti-adrenergic drugs

Alpha-2 antagonists may enhance sexual arousal by their central NA effects. One example is yohimbine, which has a modest therapeutic benefit over placebo and is generally well tolerated. Phentolamine is an alpha-1 and alpha-2 antagonist used medically to treat hypertensive crises. It has been used, in combination with papaverine, to induce erection by intra-cavernosal injection, presumably mediated by its alpha-1 antagonist effect. Phentolamine administered systemically can also improve sexual arousal, presumably by blocking both the peripheral alpha-1 receptors and enhancing central arousal by blocking the alpha-2 receptors in the brain. Some evidence of beneficial effects of phentolamine has been reported, although long-term follow-up revealed high attrition.

(c) Dopamine agonists

The main effects of dopamine agonists, such as apomorphine, are to induce genital response, probably via the hypothalamic oxytocinergic system. Early studies with apomorphine in men, while showing positive effects on erection, also demonstrated substantial side effects (mainly nausea and dizziness). Sublingual administration was developed to reduce such side effects and a number of placebo-controlled studies have shown this to improve erectile function[54] but the occurrence of side effects remains a limiting factor.

(d) Melanocortin agonists

A recent development has been bremelanotide (PT-141), a metabolite of melanotan-II, and a melanocortin analogue, which probably works through the same oxytocinergic system as apomorphine. Side effects similar to those with apomorphine occur, and further phase 3 research is required before the clinical value of this compound is established.

(i) Premature ejaculation

The orgasm inhibiting effect of SSRIs (see above) has been exploited in the treatment of PE. Paroxetine, sertraline, and fluoxetine, among others, have all been shown to be effective in this respect. However, continued use is required, and effects may not be apparent during the first week. This raises the issue of side effects and the need to avoid sudden withdrawal.

More recently, PDE-5 inhibitors have been explored as a treatment for PE. However, the benefits are not clear cut and may be restricted to cases of 'secondary' PE. The most likely mechanism of action is reduction of the smooth muscle contraction involved in seminal emission.

(ii) Delayed or absent orgasm/ejaculation

At present there is no accepted pharmacological treatment for delayed or absent ejaculation or orgasm.

(iii) Low sexual desire

The most treatable but relatively infrequent cause of loss of sexual desire in men is hypogonadism. Where androgen deficiency is evident, T replacement is indicated. Currently the most favoured route of administration is transdermal, either using cream or skin patches. One advantage of the transdermal route is that the hormone is absorbed into the skin and then released more gradually into the circulation, maintaining relatively physiological levels. In comparison, intra-muscular injections of T (e.g. T enanthate) produce supraphysiological levels in the circulation, which then steadily decline. Oral routes are complicated by first pass through the liver.

If loss of sexual desire is associated with hyperprolactinaemia, treatment with dopamine agonists, such as bromocriptine, is indicated.

Pharmacological treatments for women

At present pharmacological methods for treating sexual dysfunction in women, although generating much interest and controversy, are very limited. Most attention is currently being paid to the use of T for treating low sexual desire. As considered earlier, the evidence of a role for T in women's sexuality is inconsistent and the likelihood is that it is important for some women and not others.

The most convincing evidence of the benefits of T are in women who have been ovariectomized and have therefore experienced a substantial reduction in their androgen levels. Much of this evidence, however, indicates that supraphysiological levels of T result from treatment; thus, it remains uncertain whether the benefits result from correction of T deficiency or pharmacological effects of high levels of T. For this reason, and because of uncertainty about possible long-term risks, there has been a reluctance to approve such treatments, particularly in women with intact ovaries.[48]

Attempts to treat sexual desire/arousal problems with PDE-5 inhibitors have so far proved unsuccessful, although there may be subgroups of women who benefit. Other pharmacological approaches which are being explored in women include the use of phentolamine, apomorphine, bupropion, and bremelanotide (PT-141). But further research is required before any of these can be recommended for clinical use. For more details of the above treatments, see Bancroft.[52]

Evaluation of treatments

Sex therapy

In a 1997 review, Heiman and Meston[55] concluded that there was evidence for empirically validated treatments for orgasm problems in women and ED in men. There was inadequate support for effective treatments for low sexual desire, sexual aversions, dyspareunia in women and men, and delayed orgasm in men. What is striking is that of the 90 studies involving psychological treatments cited in this review, only two were published since 1990, and 60 of them before 1980.

In the decade since this review, there have been a few controlled trials involving cognitive behaviour therapy (CBT) for women with vaginismus,[56] vulvar vestibulitus,[57] and couples presenting with problems of low sexual desire,[58] all suggesting positive effects. Further outcome studies of psychological therapies that identify predictors of successful outcomes are badly needed.

Combined psychological and medical treatment

There have now been a number of studies in men that have compared combined psychological and pharmacological treatments, with either used separately. These have been reviewed by Althof[6] and include studies combining psychotherapy with sildenafil, and counselling with intra-cavernosal injections or with vacuum devices. In each case, the combination produced significantly better results than the medical treatment alone.

Planning a treatment programme

Assessment

When couples present with sexual difficulties, one of them is usually regarded as having the problem. However, both partners should be carefully assessed whenever possible. There are three stages to assessment; first, to facilitate the decision about whether sex therapy is appropriate; second, to identify issues relevant to the sexual problem that need to be resolved, and third, to determine whether medication or other treatments are required.

Keeping in mind the distinction between a sexual problem that is adaptive or appropriate given an individual's current circumstances and one that is maladaptive (and can perhaps be considered a 'dysfunction'), we can assess each individual's case through the three conceptual 'windows' described earlier. Are there problems in the couple's current relationship or situation, which would make inhibition of sexual responsiveness in either partner understandable or adaptive? Does either individual give a history that suggests vulnerability to sexual problems? Are there are any mental or physical health issues or medication use in either partner that could be having a negative effect on sexual interest or response?

(a) The initial interview

Although not all of the details can be obtained during the initial interview, assessment of the following topics should be carried out: the nature of the problem, including an assessment of level of sexual interest and response in each partner; identification of other assessments that may be needed (e.g. physical examination, blood tests); commitment and motivation of each partner to improving the relationship; and, assessment of the mental state of each partner. If both partners are present, each should be interviewed separately following a conjoint interview. As far as possible, questions should be asked about each individual's sexual history, the nature of the current relationship, contraceptive use and reproductive history, and alcohol or recreational drug use.

At the end of the initial interview, the clinician should provide a preliminary formulation of the nature of the problem and the types of intervention that may be helpful. Whatever treatment methods are used, it is important to continue to see the couple together to monitor progress and provide counselling as needed.

If there is no evidence of causal factors of the kind viewed through the 'first' or 'second' windows (see above), then a trial of pharmacological treatment may be appropriate. In such cases a physical examination and, where relevant, laboratory investigations would normally be arranged before starting treatment. It is important to have a good 'clinical baseline' before embarking on pharmacological interventions.

If there are any indications of problems or issues, particularly of the kind that invoke inhibition of sexual response, which need to be resolved, then pharmacological interventions should not be considered as a first step. In many cases, more assessment is required before such factors can be adequately assessed. There are then two options: (i) further interview(s) to explore such issues or (ii) starting on a programme of sex therapy. The rationale for the second option is as follows. The initial two stages of sex therapy (i.e. involving non-genital and then genital touching with no attempts at vaginal intercourse) are particularly effective at identifying relevant issues underlying the problem. Furthermore, the process involved in those early stages is likely to benefit any sexual relationship, even those without obvious problems. After three or four sex therapy sessions, a re-appraisal would be made and a decision taken whether to continue with sex therapy alone, or to combine it with an appropriate pharmacological method. For example, with a couple where the man has erectile problems, it is easier to assess the indications for the use of a PDE-5 inhibitor after the couple have gone through the first two stages of the programme where there is no 'performance pressure' and no need for an erection to occur. Similarly, it is often informative to see what impact this behavioural programme has on the individual who has complained of reduced sexual desire, before attempting to deal with the problem pharmacologically or hormonally.

Ideally, the use of the pharmacological method should then proceed in combination with the continuation of the sex therapy programme. In that way a gradual transition to a satisfying sexual relationship can be achieved without renewed 'performance pressures'. If, however, progress continues to be made with sex therapy, then the addition of pharmacological interventions may not be necessary. It should be explained to the couple that, whereas pharmacological treatments often have beneficial effects on sexual response, they do not 'cure' the problem, which is likely to recur once the medication is stopped. Furthermore, there are often side effects that have to be dealt with. There is, however, some evidence that when medication is combined with sex therapy, the pharmacological component can be gradually withdrawn, or only used intermittently.

Ethical aspects of clinical management

In assessing individuals or couples with sexual problems it is clearly important to assure them that the information they provide will be treated as highly confidential. It is appropriate to have separate case files for sexual problem clinics, rather than files used in other clinics. In this way the files can be kept secure. Some overview of the information presented to the clinic by the referring clinician should be provided at the start of the assessment. Information to be passed back to the referring clinician by letter or word of mouth should be agreed with the client(s) in advance, particularly if any highly sensitive issue emerged during the assessment. When assessing a couple, and seeing each partner separately, it is important to establish whether each individual agrees to any information not known by the other being shared during the joint sessions (e.g. details of previous sexual relationships). When offering a course of treatment or further assessment, the individual or couple should be fully informed about what is involved before being asked to make their decision.

A more challenging ethical issue relates to cultural values. Much of the process of sex therapy involves the therapist encouraging and 'giving permission' for specific forms of sexual interaction, as well as more general patterns of interaction within the relationship (e.g. 'self asserting' and 'self protecting'). It is important for the therapist to keep in mind that many of these principles are based on western middle class values. When working with a couple from a different culture, or even a different social class to the therapist, there should be awareness, openness, and discussion about contrasting values of cultural or religious origin. In this way differences can be negotiated, rather than have therapist values imposed.

When to seek specialist help

The treatment programme described above can be applied by psychiatrists in practice providing that they (i) are comfortable asking detailed questions about sexual behaviour and response, (ii) have experience working psychotherapeutically with couples, and (iii) are familiar with cognitive behavioural methods of treatment. They have the advantage over many non-medically qualified sex therapists of being able to prescribe medications in combination with a behavioural approach to sex therapy.

Specialist help should be obtained if there is evidence of a physical condition which requires detailed assessment and treatment, or if, after a reasonable trial of integrated psychological and pharmacological treatment, the sexual problem remains unresolved. In some such cases, the specialist help should be from someone with expertise in sexual medicine (e.g. urologist, gynaecologist). In other cases, where the psychological aspects remain unclear or difficult to resolve, referral to an experienced sex therapist is appropriate.

Further information

American Association of Sex Counselors, Educators, and Therapists (AASECT). http://www.aasect.org/.

British Association of Sexual and Relationship Therapists (BASRT) [http://www.basrt.org.uk].

Bancroft, J. *Human sexuality and its problems* (3rd edn). Elsevier, Oxford, in press.

Kingsberg, S., Althof, S.E., and Leiblum, S. (2002). Books helpful to patients with sexual and marital problems. *Journal of Sex & Marital Therapy*, **28**, 219–28.

Leiblum, S.R. (2007). *Principles and practice of sex therapy* (4th edn). Guilford, New York.

References

1. Masters, W.H. and Johnson, V.E. (1970). *Human sexual inadequacy*. Little Brown, Boston.
2. Kaplan, H.S. (1975). *The new sex therapy*. Bailliere Tindall, London.
3. Pridal, C.G. and LoPiccolo, J. (2000). Multielement treatment of desire disorders: integration of cognitive, behavioral, and systemic therapy. In: *Principles and practice of sex therapy* (3rd edn) (eds S.R. Leiblum and R.C. Rosen), pp. 57–81. Guilford Press, New York.
4. Althof, S. (2002). When an erection alone is not enough: biopsychosocial obstacles to lovemaking. *International Journal of Impotence Research*, **14**(Suppl. 1), S99–104.
5. Potts, A., Gavey, N., Grace, V.M., *et al.* (2003). The downside of Viagra: women's experiences and concerns. *Sociology of Health and Illness*, **25**, 697–719.
6. Althof, S. (2006). Sexual therapy in the age of pharmacotherapy. *Annual Review of Sex Research*, **17**, 116–31.
7. Rosen, R.C. (2007). Erectile dysfunction: integration of medical and psychological approaches. In *Principles and practice of sex therapy* (4th edn) (ed. S.R. Leiblum), pp. 277–312. Guilford Press, New York.
8. Laumann, E.O., Paik, A., and Rosen, R.C. (1999). Sexual dysfunction in the United States: prevalence and predictors. *The Journal of the American Medical Association*, **281**, 537–44.
9. Mercer, C.H., Fenton, K.A., Johnson, A.M., *et al.* (2003). Sexual function problems and help seeking behaviour in Britain: national probability sample survey. *British Medical Journal*, **327**, 426–7.
10. Bancroft, J. (1999). Central inhibition of sexual response in the male: a theoretical perspective. *Neuroscience and Biobehavioral Reviews*, **23**, 763–84.
11. Graham, C.A., Sanders, S.A., Milhausen, R.R., *et al.* (2004). Turning on and turning off: a focus group study of the factors that affect women's sexual arousal. *Archives of Sexual Behavior*, **33**, 527–38.
12. Laan, E. and Everaerd, W. (1995). Determinants of sexual arousal: psychophysiological theory and data. *Annual Review of Sex Research*, **6**, 32–76.
13. Basson, R. (2000). The female sexual response: a different model. *Journal of Sex & Marital Therapy*, **26**, 51–64.
14. Kinsey, A.C., Pomeroy, W.B., Martin, C.F., *et al.* (1953). *Sexual behavior in the human female*. Saunders, Philadelphia.
15. Weijmar Schultz, W., Basson, R., Binik, Y., *et al.* (2005). Women's sexual pain and its management. *The Journal of Sexual Medicine*, **2**, 301–16.

16. Binik, Y.M. (2005). Should dyspareunia be retained as a sexual dysfunction in DSM-V? A painful classification decision. *Archives of Sexual Behavior*, **34**, 11–22.
17. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, text revision) (DSM-IV). Author, Washington, DC.
18. World Health Organization. (1992). *ICD-10: International statistical classification of diseases and related health problems* (10th edn). Author, Geneva.
19. Bancroft, J., Graham, C.A., and McCord, C. (2001). Conceptualizing sexual problems in women. *Journal of Sex & Marital Therapy*, **27**, 95–103.
20. Tiefer, L. (2001). A new view of women's sexual problems: why new? why now? *The Journal of Sex Research*, **38**, 89–96.
21. The Working Group for a New View of Women's Sexual Problems. (2001). A new view of women's sexual problems. In *A new view of women's sexual problems* (eds E. Kaschak and L. Tiefer), pp. 1–8, The Haworth Press, New York.
22. West, S.L., Vinikoor, L.C., and Zolhoun, D. (2004). A systematic review of the literature on female sexual dysfunction: prevalence and predictors. *Annual Review of Sex Research*, **15**, 40–172.
23. Graham, C.A. and Bancroft, J. (2005). Assessing the prevalence of female sexual dysfunction with surveys: what is feasible? In *Women's sexual function and dysfunction: study, diagnosis and treatment* (eds I. Goldstein, C. Meston, S. Davis, and A. Traish), pp. 52–60. Taylor and Francis, London.
24. Dunn, K.M, Croft, P.R., and Hackett, G.I. (1999). Association of sexual problems with social, psychological, and physical problems in men and women: a cross sectional population survey. *Journal of Epidemiology and Community Health*, **53**, 144–8.
25. Richters, J., Grulich, A.E., de Visser, R.O., *et al.* (2003). Sexual difficulties in a representative sample of adults. *Australian and New Zealand Journal of Public Health*, **27**, 164–70.
26. Bancroft, J., Loftus, J., and Long, J.S. (2003). Distress about sex: a national survey of women in heterosexual relationships. *Archives of Sexual Behavior*, **32**, 193–8.
27. King, M., Holt, V., and Nazareth, I. (2007). Women's views of their sexual difficulties: agreement and disagreement with clinical diagnoses. *Archives of Sexual Behavior*, **36**, 281–8.
28. Feldman, H.A., Goldstein, I., Hatzichristou, D.G., *et al.* (1994). Impotence and its medical and psychosocial correlates: results of the Massachusetts Male Aging Study. *The Journal of Urology*, **161**, 54–61.
29. American Association for Retired Persons (AARP). (1999). *Modern maturity sexuality study*. NFO Research, Inc., Washington, DC.
30. Hayes, R. and Dennerstein, L. (2005). The impact of aging on sexual function and sexual dysfunction in women: a review of population based studies. *The Journal of Sexual Medicine*, **2**, 317–30.
31. Janssen, E., Vorst, H., Finn, P., *et al.* (2002). The Sexual Inhibition (SIS) and Sexual Excitation (SES) Scales: i. Measuring sexual inhibition and excitation proneness in men. *The Journal of Sex Research*, **39**, 114–26.
32. Graham, C.A., Sanders, S.A., and Milhausen, R.R. (2006). The Sexual Excitation/Sexual Inhibition Inventory for Women: psychometric properties. *Archives of Sexual Behavior*, **35**, 397–409.
33. Bancroft, J., Herbenick, D., Barnes, T., *et al.* (2005). The relevance of the dual control model to male sexual dysfunction: the Kinsey Institute/BASRT collaborative project. *Sexual & Relationship Therapy*, **20**, 13–30.
34. Varjonen, M., Santilla, P., Hoglund, M., *et al.* Genetic and environmental effects on sexual excitation and sexual inhibition in men. *The Journal of Sex Research,* **44**, 359–69.
35. Sanders, S.A., Graham, C.A., and Milhausen, R.R. (2008). Predicting sexual problems in women: relevance of sexual inhibition and sexual excitation. *Archives of Sexual Behavior*, **37**(2), 241–51
36. Lindal, E. and Stefansson, J.G. (1993). The lifetime prevalence of psychosexual dysfunction among 55 to 57-year-olds in Iceland. *Social Psychiatry and Psychiatric Epidemiology*, **28**, 91–5.
37. Osborn, M., Hawton, K., and Gath, D. (1988). Sexual dysfunction among middle-aged women in the community. *British Medical Journal*, **296**, 959–62.
38. Beck, A.T. (1967). *Depression: clinical, experimental and theoretical aspects*. Staples Press, London.
39. Segraves, R.T. (1998). Psychiatric illness and sexual function. *International Journal of Impotence Research*, **10**(Suppl. 2), S131–3.
40. Schreiner-Engel, P. and Schiavi, R.C. (1986). Lifetime psychopathology in individuals with low sexual desire. *Journal of Nervous and Mental Disease*, **174**, 646–51.
41. Ware, M.R., Emmanuel, N.P., Johnson, M.R., *et al.* (1996). Self-reported sexual dysfunctions in anxiety disorder patients. *Psychopharmacology Bulletin*, **32**, 530.
42. Angst, J. (1998). Sexual problems in healthy and depressed persons. *International Clinical Psychopharmacology*, **13**(Suppl. 6), S1–4.
43. Figueira, I., Possidente, E., Marques, C., *et al.* (2001). Sexual dysfunction: a neglected complication of panic disorder and social phobia. *Archives of Sexual Behavior*, **30**, 369–77.
44. Araujo, A.B., Durante, R., Feldman, H.A., *et al.* (1998). The relationship between depressive symptoms and male erectile dysfunction: cross-sectional results from the Massachusetts Male Aging Study. *Psychosomatic Medicine*, **60**, 458–65.
45. Bancroft, J., Janssen, E., Strong, D., *et al.* (2003). The relation between mood and sexuality in heterosexual men. *Archives of Sexual Behavior*, **32**, 217–30.
46. Lykins, A., Janssen, E., and Graham, C.A. (2006). The relationship between negative mood and sexuality in heterosexual college women. *The Journal of Sex Research*, **43**, 136–43.
47. Lilleleht, E. and Leiblum, S.R. (1993). Schizophrenia and sexuality: a critical review of the literature. *Annual Review of Sex Research*, **4**, 247–76.
48. Bancroft, J. (2005). The endocrinology of sexual arousal. Starling review. *Journal of Endocrinology*, **186**, 411–27.
49. Mustanski, B. and Bancroft, J. (2006). Sexual dysfunction. In *Psychopharmacogenetics* (eds P. Gorwood and M. Hamon), pp. 479–94. Kluwer Academic Plenum, New York.
50. Bochinski, D. and Brock, G.B. (2001). Medications affecting erectile function. In *Male sexual function: a guide to clinical management* (ed. J.J. Mulcahy), pp. 91–108. Humana, Totowa.
51. Graham, C.A., Bancroft, J., Greco, T., *et al.* (2007). Does oral contraceptive-induced reduction in free testosterone adversely affect the sexuality and mood of women? *Psychoneuroendocrinology*, **32**, 246–55.
52. Bancroft, J. *Human sexuality and its problems* (3rd edn). Elsevier, Oxford, in press.
53. Rosen, R.C. and McKenna, K.E. (2002). PDE-5 inhibition and sexual response: pharmacological and clinical outcomes. *Annual Review of Sex Research*, **13**, 36–88.
54. Von Keitz, A.T., Stroberg, P., Bukofzer, S., *et al.* (2002). A European multicentre study to evaluate the tolerability of apomorphine sublingual administered in a forced dose-escalation regimen in patients with erectile dysfunction. *British Journal of Urology International*, **89**, 409–15.
55. Heiman, J.R. and Meston, C.M. (1997). Empirically validated treatment for sexual dysfunction. *Annual Review of Sex Research*, **7**, 148–94.
56. Van Lankveld, J.D.M., ter Kuile, M.M., de Groot, H.E., *et al.* (2006). Cognitive-behavioral therapy for women with lifelong vaginismus: a randomized waiting-list controlled trial of efficacy. *Journal of Consulting and Clinical Psychology*, **74**, 168–78.
57. Bergeron, S., Binik, Y.M., Khalifé, S., *et al.* (2001). A randomized comparison of group cognitive-behavioral therapy, surface electromyographic biofeedback, and vestibulectomy in the treatment of dyspareunia resulting from vulvar vestibulitis. *Pain*, **91**, 297–306.

58. Trudel, G., Marchand, A., Ravart, M., *et al.* (2001). The effect of a cognitive-behavioral group treatment program on hypoactive sexual desire in women. *Sexual & Relationship Therapy*, **16**, 145–64.
59. Fugl-Meyer, A.R. and Fugl-Meyer, K.S. (1999). Sexual disabilities, problems and satisfaction in 18–74 year old Swedes. *Scandinavian Journal of Sexology*, **2**, 79–105.

4.11.3 The paraphilias

J. Paul Fedoroff

Clinical features

Definition of the condition

The characteristic essential to all paraphilias is the presence of a persistent and/or recurrent, sexually motivated interest that causes harm. This definition has several implications.

Interest versus act

Paraphilias can exist even if they have never been acted upon. By definition, all paraphilias begin with a sexual thought and, like non-paraphilic interests, the majority of sexual fantasies are never fulfilled in reality. Sexual acts are only paraphilic if they are motivated by harmful sexual interests. For example, an individual with paedophilia (sexual interest in children) may act on this interest by masturbating while viewing non-pornographic children's television shows. However, an individual who unintentionally downloads pictures of children from the Internet while meaning to download adult pornography should not be considered paedophilic (even though he or she may still be criminally liable).

Persistent versus opportunistic

Paraphilias are characterized by their persistence. Therefore a single paraphilic thought or activity, especially if it occurs during unusual circumstances, is unlikely to be indicative of a true paraphilia. For example, a woman who while on a vacation and under the influence of alcohol exposes herself once to group of strangers in a bar, would not normally be considered to have exhibitionism (sexual arousal from exposure to strangers). Opportunistic activity that is not sexually motivated, even if it is ongoing, is also exclusionary. A pimp who coerces women to exchange sexual activities for drugs or money, would not meet criteria for sexual sadism even though he repeatedly engages in opportunistic sexual activity with otherwise unwilling participants because, for the pimp, the motivation is financial rather that sexual.

Harm versus happenstance

Many sex acts are intimate. Therefore it should come as no surprise that participants can be harmed. Unfortunately whenever sexual activities are considered, 'harm' tends to be defined somewhat solipsistically. Paraphilias are characterized by sexual interests or behaviours in which harm is more or less inevitable rather than accidental or random. For example, although sexual intercourse may expose the participants to a number of dangers including sexually transmitted diseases, possibly unintended degrees of intimacy, or subsequent events (e.g. pregnancy), by and large, consensual sexual activity between adults does not inevitably lead to disaster.

In contrast, true paraphilic interests are by definition, harmful. For example, a woman who can only become sexually aroused if she is choked to unconsciousness (asphyxophilia), exposes herself to unintended harm including cerebral anoxia and possible death. The paraphilias associated with crime (e.g. voyeurism, exhibitionism, frotteurism, criminal sadism, and paedophilia) involve non-consensual imposition of sexual activity onto others. Other non-criminal paraphilias (e.g. transvestic fetishism) become problematic when they interfere with the ability to maintain a reciprocal emotional relationship. Most men with transvestic fetishism do not seek therapy and there is no indication they should, unless the interest begins to cause harm. Paraphilic transvestites are sufficiently dependent on cross-dressing that it causes distress. Transvestic fetishism is a good example of a paraphilic condition in which the individual symptoms (wearing women's clothing while masturbating or engaging in sexual relations) are not problematic. However, when the sexual interest becomes so pervasive that it interferes with consensual sexual relations, a diagnosis is permissible. Transvestic fetishism is very responsive to treatment with selective serotonergic re-uptake inhibitors prescribed at doses low enough to avoid inhibited orgasm.

Pathologic versus unconventional

While paraphilias are characterized by deviant sexual interests, unconventional interests alone are not sufficient to meet criteria for a true paraphilic condition. This is a persistent source of confusion in two specific situations.

(a) Homosexuality

A primary sexual interest in the same sex (homosexuality) is statistically rare. However, there is nothing about a primary same-sex interest that necessarily leads to harm to anyone. Sexual interest in a woman is no more harmful for a heterosexual man than it is for a lesbian woman. Although homosexuality is statistically and socially unconventional, the absence of inevitable or likely harm excludes same-sexed sexual interest from being paraphilic.

(b) Sadomasochism

Sexual arousal from control (sadism) or from being controlled (masochism) illustrates a second way in which unconventional sexual interests may fail to meet criteria for designation as a paraphilia. While harm is a necessary criterion for paraphilias, it is not sufficient. For example, while many conventional sports involve competition and attempted domination of an opponent, the activity is not primarily sexually motivated. Therefore, although boxing involves the intentional attempt to render an opponent unconscious via infliction of a cerebral concussion (knockout), pugilism is not a paraphilia because it is not primarily sexually motivated.

The converse is also true. Many men and women engage in interactions that are sexually motivated and which involve negotiations about power and control, domination and submission. These themes have become highly organized and regulated within social groups under the general category of 'BDSM' (bondage, domination, submission, sadism, masochism). Publications describing the BDSM 'lifestyle' universally champion the credo: 'safe, sane, and

consensual'. Therefore, men and women who engage in 'BDSM' sexual activities, although they may involve statistically and/or socially unconventional activities, do not meet criteria for any paraphilia disorder. In fact, it is arguable that anyone who is sexually aroused by the idea of engaging in 'safe, sane, and consensual' activities is *less* paraphilic than those with conventional sexual interests who are willing to compromise some of these meritorious criteria in pursuit of conventional sexual interactions. For a more complete discussion of sexual violence and sexual sadism.[1]

Classification

Table 4.11.3.1 consists of a partial list of the over 100 paraphilic disorders described in the literature. Classification of the paraphilias remains controversial. This is due to two issues. The first is that many paraphilias have been assigned names based on Latin or Greek etymology. For example, retifism refers to a paraphilic interest in shoes more commonly known as a 'shoe fetish' while 'renifleurs' are individuals with sexual arousal from the smell of urine. More complete listings of paraphilic disorders has been

Table 4.11.3.1 Paraphilic sexual disorders

Paraphilia	DSM- IV TR	ICD-10	Essential feature: persistent sexual arousal towards	Comments
Acrotomophilia	302.9	F65.9	Amputees	
Apotemnophilia	302.9	F65.9	Being an amputee	
Asphyxiophilia	302.9	F65.9	Being asphyxiated	Also known as 'autoerotic asphyxia'
Biastophilia	302.9	F65.9	Non-consensual adult intercourse	Also known as paraphilic rapism or raptophilia
Exhibitionism	302.4	F65.2	Exposure to strangers	
Fetishism	302.81	F65.0	Inanimate objects	Not vibrators
Frotteurism	302.89	F65.8	Rubbing groin without consent	ICD has no specific listing
Mysophilia	302.9	F65.9	'Filth'	Typically involving 'soiled' (worn) panties
Necrophilia	302.9	F65.9	Corpses	
Paedophilia	302.2	F65.4	Children	ICD does not differentiate
Attraction				
Males				
Females				
Both				
Exclusivity				
Incest only				
Exclusive				
Non-exclusive				
Polyembolokoilamania	302.9	F65.9	Insertion of objects	Associated with Smith McGinnis syndrome. (Not clearly paraphilic)
Scoptic syndrome	302.9	F65.9	Being castrated	
Scoptophilia	302.9	F65.9	Consensual viewing	Paraphilic only if problematic
Sexual masochism	302.83	F65.5	Loss of control	ICD combines into Sadomasochism
Sexual sadism	302.84	F65.5	Non-consensual control	ICD combines into sadomasochism
Somnophilia	302.9	F65.9	sleeping sexual partner	
Telephone scatalogia	302.9	F65.9	Obscene phone calls	
Transvestic fetishism with gender dysphoria	302.3	F65.1	Wearing clothes of the opposite sex	ICD does not subclassify
Urophilia	302.9	F65.9	Urine	
Voyeurism	302.82	F65.3	Spying	
Paraphilia NOS	302.9	F65.9	Paraphilias not otherwise specified	ICD: Disorder of sexual preference unspecified
Other disorders of sexual preference		F65.8	Other paraphilic disorders	

published (c.f. Love, 1999). In addition, many of the paraphilias involve sexual interest in the characteristics of the sexual partner(s). Often some interest in assuming the same characteristics is evident and receives a unique name. The most obvious example is sado-masochism which the DSM classification divides into sadism and masochism while in the ICD classification the two complimentary conditions are combined.

The second problematic diagnostic issue in the classification of the paraphilias concerns the need to describe both unconventional sexual behaviours and problematic sexual behaviours. For example, while sexual arousal from cross-dressing is unconventional, it technically does not meet criteria for transvestic fetishism unless it causes distress. In the case of paraphilias involving criminal interests (e.g. paedophilia) issues arise if a person reports persistent sexual interest in children but no distress or wish to act on the paedophilic interest. Newer diagnostic criteria likely will address this issue.

Diagnosis and differential diagnosis

Similar to most psychiatric conditions, paraphilic disorders are diagnosed primarily on the basis of self-reported symptoms. Paraphilias differ from most other psychiatric disorders in two ways: (i) many paraphilias involve illegal interests and (ii) objective measures of the primary criteria (in this case sexual interest) are available.

Illegal sexual interests

Paraphilic interests do not necessarily lead to illegal activities, and vice versa. Therefore, 'Not all paedophiles are child molesters and not all child molesters are paedophiles'.

With few exceptions, individuals with paraphilic interests not only know they have abnormal sexual interests but also wish they could replace them with 'normal' ones. Many confuse fantasy with reality, thinking that illegal sexual interests are equivalent to having committed a sex crime. A major issue in the diagnosis of paraphilic disorders is distinguishing between legal and psychiatric concerns (see Management section below for further details).

In addition to legal issues, clinicians should also consider several other diagnostic issues:

(a) False accusations

At one time accusations by children of sexual assault by adults were considered to be almost certainly true since it was assumed that children could not know about sex. A typical assertion would be that a child could not possibly describe acts such as sexual intercourse or ejaculation unless they had been sexually assaulted. Clearly this was before the widespread availability of pornographic videos, DVD's, cable TV networks, and the Internet.

Beginning in the 1990s, adults began to report they had been sexually abused as children but had only recently recovered their memories of the assault. In part this trend seems to have been due to two factors: the decision in the United States to reset the time at which the statute of limitations required a sexual offence to be reported to the time at which the offence was recalled; and the believe that failure to recall sexual abuse was a sign that it had occurred.

(b) False confessions

While less frequent, false reports of paraphilic interests have also been described.[2] The most frequent presentation of false paraphilic symptoms takes the form of a man or woman with depression who reports obsessions involving often exceptionally troubling sex crimes. While a detailed phenomenologic examination of this phenomenon has not been published, several characteristics are typical. The individual often has a history of a mood disorder or is in circumstances in which affective disorders are more likely (e.g. post-partum). The sexual obsession typically involves children to whom the patient has access (it is rare for the patient to report spontaneous fantasies of sexual interactions with unknown children). Most importantly, the fantasies are extremely ego-dystonic. Asked if they ever masturbate to their paraphilic fantasies, they typically respond with horror and, unlike those with true paraphilias, often describe self-loathing indicative of a change in self-esteem due to depression. A danger of false confessions or admission of false paraphilic interests has also been noted in men and women with intellectual disability.

(c) Co-morbid conditions

People rarely seek treatment on their own specifically for paraphilic disorders. This is due to unfortunately widespread false beliefs that (i) there are no effective treatments for paraphilias, (ii) embarrassment about discussing paraphilic symptoms, and (iii) in the case of paedophilia, the mistaken belief that reporting a sexual interest in children necessarily requires that the patient be reported to the police (please see Management below for more comments on this problem). Since paraphilias themselves rarely motivate helpseeking, clinicians should include other conditions in the differential diagnoses both as alternative explanations for the problem and as co-morbid conditions that may be present in addition to the paraphilia. The most frequent of these are mood disorders, substance abuse problems including alcohol, marital disorder, and legal problems.

Less common are organic disorders including brain injuries.[3] Surprisingly, given the importance of sex hormones in the development and expression of sexual characteristics, endocrine disorders affecting the sex hormones are rarely implicated. This may be because testosterone in men with normal or elevated hormone levels are more closely associated with violence and aggression than with alterations in the direction of sexual interest.[1] One exception is hypogonadism associated with Klinefelter's syndrome. In some men with Klinefelter's syndrome, paraphilic problems become apparent when testosterone is prescribed to correct hypogonadism. In those men with Klinefelter's syndrome and paraphilic interests, addition of testosterone appears to unmask rather than cause previously unexpressed paraphilic conditions.

A more controversial question involves a possible association between paraphilias and intellectual disability. Some research has supported the hypothesis that, as a group, men with paedophilia have below average intelligence. However, alternative explanations are possible. For example, most men with paedophilia come to attention when they are arrested. The fact that men with intellectual disability are more likely to be arrested and are over-represented in the criminal justice system may skew the data. Those who live in supervised housing are also more likely to have private activities not only discovered by staff but also labelled by staff as deviant. In addition, men with intellectual disability frequently have impairments in social skills that can certainly contribute to problematic sexual behaviours independent of paraphilic interests. This phenomenon has been described as 'counterfeit deviance'.[4]

One meta-analytic study supporting an association between paedophilia and intellectual disability found the mean I.Q. of sex offenders to be only five I.Q. points below the mean I.Q. of non-sexual offenders.[5] While cognitive ability is important in determining level of risk and in planning treatment (see below) it would be a diagnostic mistake to confirm or refute a diagnosis of paedophilia on the basis of intelligence.

A further area of controversy involves the question of whether having one paraphilia predisposes to having other paraphilias? The answer depends on whether the assessor is a 'lumper' or a 'splitter'. John Money viewed paraphilias as 'vandalized lovemaps' that were unique. He argued that a paedophile might begin by spying on children, then surreptitiously touching them, then engaging in sexual relations. It was his view that it was more accurate to label the disorder as paedophilia (since this explains the motivation behind the varied behaviours) as opposed to making a diagnosis of voyeurism, frotteurism, and paedophilia. Clearly both approaches have strengths and weaknesses. Most important is to be aware of both diagnostic methods when evaluating incidence or prevalence reports.

Epidemiology

Prevalence of paraphilic disorders

Any discussion of the number of people with paraphilic disorders in the population must begin with a series of caveats. The most important is the fact that the majority of the information available is derived from studies of men convicted of sex crimes. This is a significant problem since not all paraphilias are associated with sex crimes. The problem is compounded by the fact that many paraphilic disorders remain undiagnosed either because assessment was never requested, clinicians did not gather sufficient information, or because the person with the paraphilic disorder was unwilling or unable to disclose the symptoms. Further degrees of confusion and subsequent dispute are added by inconsistent application of existing diagnostic criteria (for example confusing child molesters with paedophiles) or by differing opinions about whether or not to subdivide paraphilic disorders (e.g. diagnosing a person who lures children into sexual activity by exposing to them as a paedophile or as both a paedophile and an exhibitionist). To date, insufficient attention has been paid to the importance of precise definitions of what is being measured, the difference between point and period prevalence, and the potentially significant differences attributable to independent characteristics of the populations being studied. For example, a report on the incidence of 'sexual sadism' based on a point prevalence study of sexual assault of women and children in a country in which war is being waged, while important, has little to do with an analysis of the period prevalence of sexual sadism in, for example, North America. There is also an widespread but unwarranted assumption that studies based on criminal populations can be easily generalized to other populations.

Fortunately, with these above caveats in mind, a significant resource to answer epidemiologic questions is now available both in text form and, more importantly, in the form of a constantly updated Internet Web site: http://www.kinseyinstitute.org/ccies/. This Web site, currently under development and hosted on the Kinsey Institute web page, provides free access to reports by noted sexologists on sexual behaviours in 60 countries and six continents. Included in the chapters on most countries are sections dealing with both paraphilic disorders and 'unconventional sexual behaviours'. This is important since different cultures may place different significance on sexual behaviours both in public and private. The fact that each chapter is multi-authored by sexologists who live in the country being described also likely adds to the credibility of the information.

From perusal of the information available several statements about the epidemiology of paraphilic disorders can be made with reasonable confidence.

(a) Men are more likely than women to be arrested for sexually motivated crimes

Most (but not all) illegal activities are non-consensual. Those that are sexually motivated have a high likelihood of being associated with a paraphilia. Data to support this observation is derived primarily from review of the unequal ratio of men and women convicted of sex crimes. While there are increasing numbers of women (especially adolescent females) being convicted of child sexual abuse, men are still the vast majority. From a criminologic perspective men seem more willing to commit crimes of all types, especially those that involve confrontation (verbal, physical, and sexual). This is of significance since two widely employed actuarial risk assessment instruments, the Violence Risk Appraisal Guide (VRAG) and Sex Offender Risk Appraisal Guide (SORAG) rely heavily on the Hare Psychopathy Checklist (HPCL) which itself has been associated with criminal activity independent of sexual interest. This may explain why some researchers believe that a combination of high scores on the HPCL combined with deviant scores on phallometric testing are associated with high risk of re-offence (either violent and sexual) (for more information please see the Hare Web site listed below).

(b) With the exception of internet-related sex crimes, the frequency of sex crimes of all types are on the decline

This welcome finding has been reviewed extensively elsewhere.[6] It is important to note that while this trend of decreasing frequencies of sex crimes excludes countries in the midst of war or major political and social upheaval, the trend is occurring worldwide. Unfortunately the reason for this trend is unknown. One controversial explanation is a wider availability of pornographic materials, beginning with videotapes in the 1980s and now exploding with the Internet. One study noted, during a time of increased availability of pornographic materials, an 85 per cent decrease in numbers of juvenile offenders in Japan from 1803 in 1972 to 264 in 1995.[7] The increase in Internet-related crimes combined with an apparent decrease in other types of sex crimes invites the unproven speculation that one may be substituting for the other.

(c) Paraphilic disorders are not new

A final important point to be made about the epidemiology of paraphilic disorders is that they are not new. Historic texts on the topic clearly describe not only sex crimes but also the same phenomenologic characteristics seen by clinicians today. If there is a change, it is towards increasingly earlier detection of paraphilic behaviours, not only at a younger age,[8] but also before criminal acts have been committed.

Table 4.11.3.2 summarizes generally accepted prevalence estimates for the DSM-IVTR paraphilias.

Table 4.11.3.2 Frequency/prevalance of paraphilias

Paraphilia	Prevalence/frequency	Comments
Exhibitionism	No reliable data	40–60% of females report having been exposed to Evidence suggests a decrease in frequency of exhibitionism after age 40
Fetishism	No reliable data	Depending on population studied, rates range from 0.8% to 18% of men
Frotteurism	No reliable general population data	Only paraphilia claimed to decrease in frequency after age 25 (DSM-IV-TR p. 570)
Paedophilia	Frequency: 300 000 abused children per year in the USA	1988 data; does not account for repeat offences
Sexual sadism	3% to 20% of general population	Data do not distinguish between criminal and non-criminal sexual interests
Sexual masochism	5% to 10% of general population	Data does not distinguish between criminal and non-criminal sexual interests
Transvestic fetishism	1% of men	True prevalence unknown in part due to idiosyncratic diagnostic criteria
Voyeurism	No reliable data	Most studies of voyeurism involve offenders with co-morbid paraphilic interests

Aetiology

Like most psychiatric disorders, the ultimate causes of paraphilic disorders are unknown. Like most groups of psychiatric disorders, they undoubtedly have multiple contributing factors.

Four major explanatory perspectives have been identified: 'disease', 'behavioural', 'dimensional', and 'life story'.[9] Each perspective has had advocates that can be traced to the beginning of the twentieth century.

The disease perspective

The disease perspective is perhaps the most familiar to physicians since it is based upon an attempt to combine signs and symptoms into syndromes or disorders via pathophysiologic mechanisms.

In 1886 Krafft-Ebing published *Psychopathia Sexualis: A Medico-forensic Study* in which he described a large series of patients with a variety of sexual disorders, whom he had examined in his forensic psychiatry practice.[10] Using a disease perspective he advanced the hypothesis that masturbation caused a physiologic imbalance that was a major causal factor in development of sexual deviancy. Krafft-Ebing was widely criticized for his acceptance and perpetuation of the degeneration theory of masturbation (the now discarded theory that masturbation itself could cause progressive degrees of physiologic harm leading to increasingly severe psychopathology), and for his failure to clearly differentiate between extreme and mild forms of sexual deviancy.

While *Psychopathia Sexualis* is now considered a landmark text in human sexuality, its acceptance was limited because it was published at a time when other etiologic explanations for the paraphilias were becoming important. One criticism of Krafft-Ebing's disease perspective was his tendency to focus on illness rather than wellness or normality. In this context, Brautigam's criticism of Krafft-Ebing's *Psychopathia Sexualis* foreshadows current criticisms of etiologic theories of the paraphilias that rely on the disease perspective:

> This first great and distinguished complete presentation and inventory of that field has the character of a large catalogue of perversions. In a picture book-like series are offered the most monstrous cases of his time and history, particularly cases collected by Krafft-Ebing himself as a widely sought medico-legal consultant. The perversions are described in their most extreme forms. This collection of brutal necrophiliacs, anthropophages, sexual murderers,and sodomites, of cunning and cultured sadists, masochists, and transvestites has in its degree of deviation never been surpassed. By this extreme presentation Krafft-Ebing has moved sexual disorders away from general sexuality.[11]

Magnus Hirschfield was a contemporary of Krafft-Ebing who also employed a disease perspective in an attempt to identify abnormal pathophysiology in individuals with identified sexual problems. Although he is best known for his studies of castrated males and for his emphasis on the importance of hormones on sexual behaviour, Hirschfield also drew an important distinction between the simple description of observed behaviours and proof of their aetiology. His argument that describing a behaviour was not sufficient to explain its cause led him to distinguish between homosexuality, transsexualism, and fetishistic transvestism, in which the same behaviour (cross-dressing) could be shown to have different motivations and therefore presumably different causes.[12] His observation that sexual orientation and behaviour may have variable determinants was important since it emphasized the need to look for underlying causes of behaviour from a disease perspective while at the same time opening the door for collaboration with behavioural researchers.

Behavioural perspective

Although Albert Moll felt that sexual disorders could be treated with 'association therapy' (a precursor to modern behaviour modification therapy), he felt learning alone could not account for all sexual deviations:

> Were a single sexual experience, and, indeed, the first sexual experience to induce a lasting association between sex drive and the object of the first sexual experience, then we would have to find sexual perversion everywhere. Where are there to be found people who initially satisfied their sexual impulse in a normal manner?[13]

By extending consideration of sexual development from childhood to adulthood, Moll was able to answer Brautigam's criticism of Krafft-Ebing's research by studying normal as well as abnormal sexuality.

Albert Moll and Alfred Binet, used behavioural perspectives to study sexuality by directing attention to the associations between behaviours and their reinforcers. Binet re-examined the case histories of Krafft-Ebing's patients and argued that sexual deviations which had been presumed to be caused by disease or degeneration could have been caused by learned associations. He noted that many of the reported cases he studied also had 'chance associations' that could explain the development of the sexual deviation. These observations implied that, given a specific set of reinforced

experiences, any individual could become paraphilic. The proposition that all people could potentially acquire paraphilias if they were unfortunate enough to have been exposed to the necessary abnormal formative experiences was more fully developed by investigators using a dimensional perspective.

Dimensional perspective

The dimensional perspective is characterized by an avoidance of rigid categorical classifications in favour of continuous descriptors. Variations in sexual interests, including paraphilic interests, are explained as statistical extremes of normal behaviour without necessarily implying that a disease process is present. From a dimensional perspective, paedophilia would be understood as an extreme variation of the widespread tendency to equate youth with sexual attractiveness. The work of Haveloch Ellis is a good example of this approach to sex research. He is most well known for his six volume *Studies in The Psychology of Sex*.[14] A major theme in Ellis' work is the avoidance of a disease perspective as the sole explanation for sexual deviations. Instead of paraphilic disorders, he focused primarily on homosexuality and suggested that homosexuality could be 'latent'. Since degree of 'latency' is a continuous variable it was possible to consider homosexuality as a dimension rather than a disease. This dimensional perspective was later exploited by Kinsey[15] who developed his famous Kinsey scale of sexual orientation that rated men and women on a six point scale. Both Ellis and Kinsey shifted attention from the differences between homosexuals and heterosexuals to their similarities. In this way their work complemented research using disease or behavioural perspectives as well as the work of investigators using a fourth perspective, which focused primarily on the individual's 'life story'.

Life-story perspective

The life-story perspective is characterized by a search for the 'meaning' behind behaviour. This perspective appeals to the self-experienced aspect of life, so life-story perspectives tend to be very compelling and persuasive. For example, in North America, it is rare to find a man with transvestic fetishism that does not recall being dressed as a girl for Halloween. A problem with this approach is that it is often difficult to determine which 'story' is the correct one, or even if a single meaningful connection can be found. For example, there is no evidence that men with transvestic fetishism are more likely to have been dressed as girls when they were children than were other men. Sigmund Freud often used sexual motives to construct a life story that was difficult to disprove. Freud's research method consisted chiefly of interviewing patients and interpreting recurring themes in literature and art. While this approach may seem highly susceptible to bias, it has been surprisingly successful, not only for Freud but for other researchers who have adopted this method of direct interview combined with longitudinal follow-up, to understand the meanings of sexual acts and decisions. The explanations that individuals themselves develop to explain their behaviours are often as important and enlightening therapeutically as their 'true' causes.

Integration of the four perspectives

While each of the four described perspectives provide a unique basis on which to base explanatory theories about the origins of paraphilic disorders, none is mutually exclusive. In the twentieth century, some behaviourists began to acknowledge the importance of non-observable factors such as 'instinct'[16] and later, 'motivation'. Other theorists,[17] while recognizing the importance of behaviourism, emphasized that the 'motivated behaviours' of sleeping, eating, and mating could not reasonably be reduced to simple stimulus response associations. If motivated behaviours could not be fully explained from a single research perspective perhaps an integration of approaches would be more helpful. A good example of an integrated etiologic theory to explain paraphilias is the one advanced by John Money who theorized that all children are born with the potential to develop a full range of sexual interests analogous to the observation that all healthy children at the time of birth are capable of learning to speak any human language. Which language (and by analogy which sexual interests) each child eventually expresses is determined by learning during critical periods of neurologic development.

From extensive research involving patients with genetic and endocrine disorders that interfered with normal anatomic and physiologic sexual development, he theorized that sometime before the onset of puberty, a 'lovemap' was established. He defined lovemap as: 'a developmental representation or template, synchronously functional in the mind and the brain, depicting the idealized lover, the idealized love affair, and the idealized programme of sexuoerotic activity with that lover, projected in imagery and ideation, or in actual performance'.[18] He hypothesized that an individual's lovemap became 'neurologically embedded' or 'imprinted', analogous to the way an individual's native language becomes a 'native tongue'. He believed that each person has a single lovemap (analogous to having a single native language or 'mother tongue') but that each person may never be fully aware of the true nature and details of the lovemap (analogous to not learning the full vocabulary of grammar of a native language). Paraphilias represented what he termed 'vandalized' lovemaps resulting from an abnormal experience during the 'critical period' during which lovemaps are formed.

Course and prognosis

While the incidence of paraphilias (especially those that involve criminal activities) in men is at least 10 times the incidence in women,[19] the majority of both males and females do not develop paraphilias. In addition, unlike sexual orientation, for which there is some evidence of important genetic or at least prenatal influences,[20] and excluding a single retrospective report of possible familial co-morbidity,[21] there is no current evidence of any prenatal factors that cause paraphilic disorders. Similarly the hypothesis that being the victim of sexual abuse predisposes victims to become sex offenders has also been challenged.[22] There is an increasing recognition of the fact that adolescents are capable of committing sex crimes.[8] This edited book concluded that, 'approximately 20 per cent of all rapes and between 30 per cent and 50 per cent of child molestations are perpetrated by adolescent males'. The increasing number of facilities designed to assess and treat juvenile sex offenders, male and female, supports self-report descriptions of paraphilic behaviours beginning in adolescence.

Most crimes (both sexual and non-sexual) are committed by young adult males. One problem in assessing the literature on the frequency of paraphilic behaviours in this age group is the frequent failure to distinguish between current paraphilic acts

and those that are recalled. For example, it is not uncommon for a middle-aged man with exhibitionism to recall frequent exhibitionistic activities from his youth[2] but there is no reason to believe that the recollections of the sexual exploits of exhibitionists are any more reliable than the recollections of non-paraphilic men.

To summarize, it is often assumed that anyone who has a paraphilic interest has acted on it. In fact, the DSM-IV diagnostic criteria B requires that the interest has been acted upon or at least caused harm to someone. While cases have been described in which a person engaged in sexual activity without awareness (e.g. engaging in sexual acts while asleep),[23] for the most part, thought precedes action. In fact, it is arguable that a person who fantasizes about paraphilic activities is more paraphilic than one who commits a paraphilic act while fantasizing about a non-paraphilic activity.

Most paraphilic thoughts begin around the time of puberty. If the 'vandalized lovemap' explanation of paraphilias is correct, faulty development of sexual interest likely occurs prior to puberty but is not expressed until there is sufficient increase in sex drive for it to be manifested, first as fantasy and later, by action.

Prognosis is an area of considerable controversy. However, accumulated evidence indicates that the majority of convicted sex offenders do not re-offend (likely fewer that 15 per cent). In addition, evidence indicates that treatments are improving in efficacy (see below).

Evaluation of treatments

In this section 'treatment' refers to intervention with intent to reduce the frequency of unwanted paraphilic symptoms. Studies on the treatment of paraphilic disorders necessarily include participants with diagnosed paraphilias who have not only sought treatment but who have also volunteered for a treatment and have at least started the treatment under study. This represents a group that for obvious reasons may be unrepresentative of the true paraphilic population, the majority of whom likely never seek treatment.

Treatment recommendations should be individualized across paraphilic and co-morbid disorders and to account for individual variation. For example, the treatment approach for an exhibitionist with anti-social personality disorder and heroin dependency will be very different from the treatment of a woman with paedophilia or an adolescent with intellectual disability and transvestic fetishism.

In the past half-century there have been three major innovations in the treatment of paraphilic disorders that have led to significant improvement in prognosis. The first of these was the use of anti-androgen medications.

Interventions to reduce testosterone

Surgical castration has shown a high degree of success, at least in terms of reducing recidivism rates in sex offenders, with reported rates of recidivism after 30 years of follow-up as low as 2.2 per cent for convicted rapists. While this is a remarkable success rate, surgery for this purpose has been criticized on ethical grounds and because reversible interventions, in the form of anti-androgen medications have became available. For a review of the efficacy of castration in the treatment of the sex offender.[24]

Anti-androgens administered either by mouth or by intra-muscular injection were first used in the late 1950s as an alternative to surgical castration. The most common oral anti-androgens prescribed for this purpose are cyproterone acetate (Androcur) (CPA) and medroy-progesterone acetate (Provera) (MPA). Both of these medications are also available in a formulation suitable for intra-muscular injection. Reviews of treatment efficacy of both of these medications are also available.[24]

A review of eight studies involving a total of 452 sex offenders who received treatment with MPA for between 1 and 13 years, recidivism ranged from 1 per cent to 17 per cent.[25] Unfortunately, anti-androgen treatment of sex offenders has been a victim of its own success. Because the success rates were so high in open studies, and because the consequences of recidivism are so great, randomized double-blind treatment studies have been difficult to justify from an ethical perspective. Gonadotropin hormone releasing hormone (GnRH) partial analogs significantly reduce serum testosterone by down-regulating receptors in the anterior pituitary resulting in a decrease in luteinizing hormone which in turn decreases production of testosterone by leydig cells in the testes. Reviews of the apparent impressive efficacy of anti-androgens and GnRH analogs are available.[26]

Selective serotonin re-uptake inhibitors (SSRIs)

Numerous reports have confirmed the apparent efficacy of SSRI and related medications in the treatment of men with paraphilic disorders.[27] Two important prescribing issues are important:

(a) There is no evidence that the efficacy of SSRIs in the treatment of paraphilic disorders is due to suppression of sex drive. Many medications that are more effective in reducing sex drive appear to have little effect on paraphilic interest.

(b) There is no evidence that higher doses of SSRIs are more effective than low doses. In fact, if the prescribed SSRI causes inhibited orgasm, this may be counter-therapeutic since there is a tendency for individuals with paraphilic sexual interests to resort to paraphilic fantasies or acts in order to facilitate orgasm.[28,29]

Psychotherapy

The evidence in support and against various psychotherapeutic interventions has recently been reviewed.[30] Most reviews of psychotherapeutic interventions for paraphilic disorders focus on group therapy (the most common intervention in prisons) or individual psychotherapy (most common in private practices). However, other interventions are also important including marital therapy, family therapy, spousal therapy, occupational therapy, and substance abuse therapy.

Treatment of sex offenders based on psychologic interventions were carefully reviewed by a collaborative group using meta-analytic techniques.[31] A total of 43 published studies involving a total of 9454 sex offenders (5078 treated and 4376 untreated) were analysed. While noting methodologic limitations, there were two important findings: (i) sexual recidivism rates were lower in treated groups (12.3 per cent) than in comparison groups (16.8 per cent) and (ii) efficacy of psychologic treatments appear to have improved since the 1970s. The final results of one important study were not included in the Hanson *et al.* meta-analysis. This was a longitudinal study with stratified randomization into a group that received

relapse-prevention group therapy ($n = 167$) or a group that did not ($n = 225$); a third group was matched to the first two groups and consisted of sex offenders who did not volunteer to participate in the study ($n = 220$). Treatment for the first group consisted of 2 years in custody, group relapse-prevention therapy followed by a 1 year mandated programme after release to the community. While results were complicated by a dropout rate in the treatment group of 27 per cent, the third group consisting of 'non-volunteer' controls was found to have the lowest long-term rate of sexual recidivism (19.1 per cent at 12 years follow-up). In comparison, the sub-group that entered therapy but discontinued prematurely had a 35.7 per cent sexual re-offence rate compared to the group that completed the prescribed treatment which had a 21.6 per cent sexual re-offence rate at 12 years follow-up. The final conclusion of this study was that a treatment effect for the cognitive behavioural programme did not produce a significant treatment effect.[32]

The failure to demonstrate the efficacy of 'relapse-prevention' as it was constituted 20 years ago has been regarded by some as a major setback in the field. However, the fact that even in the worst outcome group the majority of sex offenders were not known to have re-offended, the fact that the treatment programme under investigation was 'manualized', and the fact that other important treatment interventions were not employed, makes any conclusion that sex offenders can not be treated unwarranted.

Management

General principles

Summaries of management strategies for individuals with paraphilic disorders and/or sex offence histories are available.[30,33] Specific recommendations for the more commonly encountered paraphilic disorders and for specific paraphilic disorders should also be consulted.[2] An initial approach to the management of paraphilic disorders is summarized in Fig. 4.11.3.1.

Individuals with paraphilic disorders typically are in extreme crisis when they present for treatment. It is important to assist the individual to not only understand that sexual urges are controllable but that sex acts are the person's responsibility. This is often a surprise to people with paraphilias who have lived with the false belief that the difference between themselves and the rest of the world is a lack of will power. It is important to assist in maintaining and establishing a healthy social support system, ideally including family members, spouses, employers, church members, and trusted friends. Group therapy is particularly helpful in assisting to confront cognitive distortions and in demonstrating that others not only do not re-offend but are capable of establishing healthy and fulfilling lives. Medications can be prescribed with the aim of decreasing sexual anxiety and impulsivity (SSRIs), moderating sex drive (anti-androgens), or eliminating sex drive (GNRH

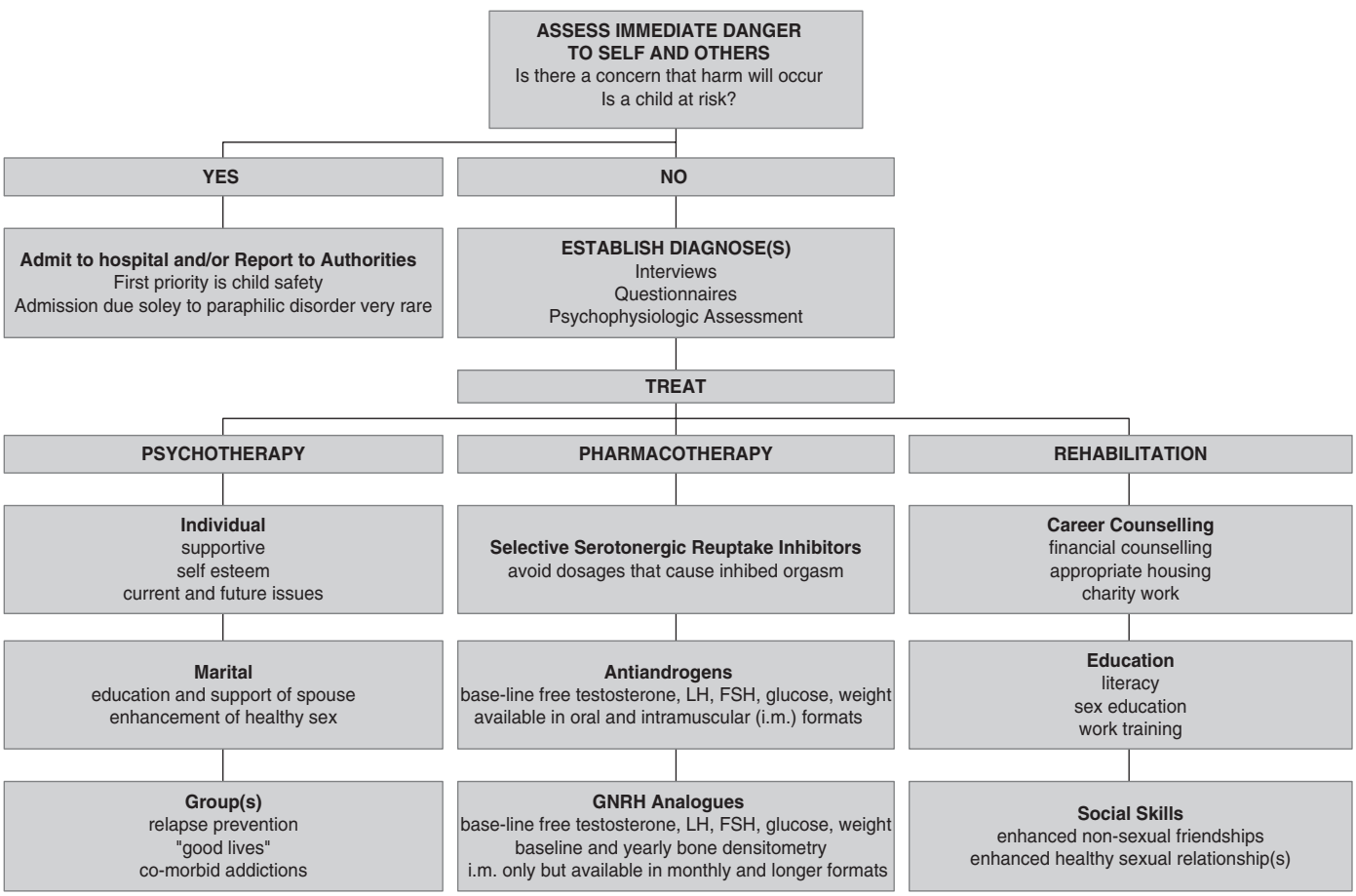


Fig. 4.11.3.1 Management of the paraphilias.

analogues). It is vital, particularly when treating men with paraphilias who have been incarcerated, to assist them in establishing pro-social lifestyles. This includes advice about appropriate work (e.g. avoiding jobs that require contact with children in the case of paedophiles), finding work (some clinics maintain lists of employers willing to hire men they know are in treatment), and ways to spend leisure time in constructive and pro-social ways. A part of the therapy often neglected is the establishment of healthy sexual activities to replace the problematic sexual interests and acts for which therapy was initiated. This frequently involves marital or couple's therapy. Some clinics establish 'spouses' groups to provide independent support, education, and treatment of the spouses of individuals with paraphilias.

In addition, the following general principles are recommended:

(a) Identify for whom you are working

Many paraphilic disorders predispose to unconventional or criminal activities. Most countries now have mandatory reporting requirements for incidents of known or potential sexual abuse of any identifiable child. It is extremely important to establish the limits of confidentiality prior to or at the first meeting. Many clinics post a description of the rules of confidentiality under which they operate. Sometimes it is difficult to be certain for whom one is working. A quick rule of thumb is to ask who is paying for the work and to whom reports will be sent.

(b) Facilitate disclosure

Men and women seeking assessment and treatment for paraphilic disorders almost always think they are the first (and worst) case of their type ever seen. They fear they will be led from the clinic room in handcuffs and be featured on the front page of the following day's newspaper. It is important to dispel these misconceptions not only because it will allow for a more in-depth assessment but also because it is important to dispel a 'me-against-authority' mentality which is common amongst individuals who see themselves as 'outsiders'.

Objective measures of sexual preference (e.g. phallometric testing) are available.[34] Their utility is enhanced if they are used as 'truth facilitators' rather than as 'lie detectors'.[35] While tests based on deception such as surreptitious measurement of time spent viewing pictures are currently widely used, their future utility is questionable since they are valid only if the deception is effective.

Perhaps the greatest truth facilitator is scrupulous adherence to disclosed rules of confidentiality.

(c) Establish what the problem is

Typically, individuals with paraphilic disorders present in crisis. However, the crisis often has only peripheral connection to the paraphilia. For example, a woman may have sexual masochism but seek treatment due to complications arising from major depression or substance abuse. This is true even if she insists that all her problems would end if she could be cured of her masochistic interests. As reviewed above, paraphilic disorders often occur together with other psychiatric, medical, and relationship problems. Clinicians should be wary of limited differential diagnoses, especially those that attribute all problems to a single paraphilia in isolation.

(d) Identify why the patient is seeking treatment now

Most paraphilias begin in childhood or at the time of puberty. However, teenagers rarely self-present for assessment of paraphilic disorders. Therefore, most adults with paraphilic disorders may be assumed to be seeking treatment for a variety of reasons besides having suddenly discovered they have (for example) paedophilia. Frequent motivations include: criminal charges, discovery of problematic behaviours by spouses or family members, and emergence of co-morbid problems such as depression or substance abuse.

For treatment to be successful it is very important to establish what the person seeking treatment thinks is the problem. Failure to do so significantly reduces the likelihood of keeping the person in therapy, establishing a therapeutic relationship, and ultimately the prognosis for a successful outcome.

(e) Avoid one-sided treatment plans

Paraphilic disorders by definition involve an interaction between the most abstract and 'high-level' cognitive processes (sexual fantasies), limbic and sub-cortical neural pathways (sexual arousal patterns), and society. Treatment plans based on single perspectives are unlikely to succeed. Most importantly, it is vital to pay attention to the worldview of the person seeking treatment. While the theory they have may be incorrect, ignoring it is likely to lead to frustration.

(f) Intervene quickly

A common mistake is to assume that because a paraphilia has been present for a long time, that it will take a long time to alleviate the symptoms. Many individuals with paraphilias think they must first review their childhood and all their previous offences before they can think about altering their current problematic behaviours. This is false. On the basis of current evidence, treatment should focus on current and future behaviours with an expectation that sexual behaviours of all types are under voluntary control. Requests to 'taper' problematic sexual behaviours should be dealt with as intentional or unintentional efforts to delay effective intervention.

(g) Be persistent

Successful treatment is not only multi-faceted but above all persistent. A useful strategy is to present treatment as a series of 'experiments' based on feedback about the effectiveness of interventions. Treatment should be presented with the expectation of success. A review of the fact that the majority of outcome studies find that the likelihood of sex offenders to recidivate is less that the likelihood of patients with major depression to have a recurrence of their life-threatening illness, is often an inspiring surprise to people entering therapy.

(h) Be inclusive with treatment

Begin with a thorough assessment and consideration of a complete differential diagnosis. If there are indications of co-morbid problems (e.g. foetal alcohol syndrome, genetic disorders, mood disorders, psychotic disorders, personality disorders, etc.) further investigations and interventions should proceed without delay. Once a diagnosis has been made, all the treatment options should be reviewed with the person seeking treatment (see above). Risks and benefits of treatment options (including the risk of declining treatment) should be thoroughly reviewed. Choice of treatment should be a collaboration between the clinician and person seeking treatment, but ultimately should be informed and voluntary. Typically treatment should include both individual and group psychotherapy. If the person has a sexual partner, couple's therapy is often of assistance. Pharmacologic treatment can include SSRIs

(often more effective at doses used for depression as opposed to the higher doses used for obsessive–compulsive disorder), anti-androgens and/or GnRH analogues. Acceptance of treatment with anti-androgens is increased by explaining that treatment effects are reversible. More information about treatment options is available.[2]

(i) Be optimistic

The frequency of sex offences is dropping. The effectiveness of treatment of sex offenders (who are not necessarily paraphilic) is improving. Sex offender re-offence rates are below 15 per cent. New as well as established treatments are available.

Beyond eliminating criminal and harmful sexual behaviours, a key to successful therapy is the establishment of non-criminal and beneficial sexual interests and behaviours. Remarkably, the literature on treatment of paraphilic disorders almost completely ignores the concept of establishing non-paraphilic sexual behaviours as a treatment goal. This is a mistake. Adult men and women rarely want to 'give up' sex. However, if they are provided with the opportunity to improve their sex lives they frequently become enthusiastic (and successful) participants.

Acknowledgements

Dr Fedoroff gratefully acknowledges Drs J. Bancroft, F. Berlin, J. Bradford, E. Coleman, P. Fagin, R. Rosen, and the late J. Money for their significant influences on his thinking about the paraphilias. However, none should be held responsible for any of the opinions expressed in this chapter.

Further information

Web sites

Robert Hare Psychopathy Web site: http://www.hare.org/

Magnus Hirschfeld Archive for Sexology: http://www2.hu-berlin.de/sexology/

Kinsey institute: http://www.indiana.edu/~kinsey/

Useful summaries

Bancroft, J. (in press). *Human sexuality and its problems*. Elsevier Health Sciences In Press.

Love, B. (1999). *Encyclopedia of unusual sex practices*. Greenwich Editions, London.

Money, J. (1986). *Venuses penuses*. Prometheus Books, Buffalo, New York.

Francoeur, R.T. and Noonan, R.J. (eds.) (2004). *The continuum complete international encyclopaedia of sexuality*. Continuum, New York.

Schwartz, B.K. (ed.) (1995–2005). *The sex offender* (Vols. 1–5). Kingston, New Jersey.

References

1. Fedoroff, J.P. (in press). Sadism, sadomasochism, sex and violence. *Canadian Journal of Psychiatry*.
2. Fedoroff, J.P. (2003). Paraphilic worlds. In *Handbook of clinical sexuality for metal health professionals* (ed. S.B. Levine). Brunner/Routeledge, New York.
3. Coleman, E. (2005). Neuroanatomical and neurotransmitter dysfunction and compulsive sexual behavior. In *Biological substrates of human sexuality* (ed. J.S. Hyde), pp. 147–69. American Psychological Association, Washington, DC.
4. Hingsburger, D., Griffiths, D., and Quinsey, V. (1991). Detecting counterfeit deviance: differentiating sexual deviance from sexual inappropriateness. *Habilitative Mental Healthcare Newsletter*, 51–4.
5. Cantor, J.M., Blanchard, R., Robichaud, L.K., *et al.* (2005). Quantitative reanalysis of aggregate data on IQ in sexual offenders. *Psychological Bulletin*, **131**, 555–68.
6. Lalumière, M., Harris, G.T., Quinsey, V., *et al.* (2005). *The causes of rape*. American Psychological Press, Washington, DC.
7. Diamond, M. and Uchiyama, A. (1999). Pornography, rape and sex crimes in Japan. *International Journal of Law and Psychiatry*, **22**, 1–22.
8. Barbaree, H. and Marshall, W.L., (eds.) (2005). *The juvenile sex offender* (2nd edn). Guilford Press, New York.
9. McHugh, P.R. and Slavney, P.R. (1983). *The perspectives of psychiatry*. Johns Hopkins University Press, Baltimore.
10. Krafft-Ebing, R.V. (1906). *Psychopathia sexualis. A medico-forensic study, with especial reference to the antipathic sexual instinct*. Physicians and Surgeons Book Company, New York.
11. Brautigam, W. (1977). Die angliche Beurtelung in Psychopathologie der Sexualitae. In (eds. H. Grese and F. Stutgart).
12. Hirshfield, M. (1910). Die Transvestiten. A. Pulvermacher, Berlin.
13. Moll, A. (1983). Untersuchungen uber die Libido Sexualis. In *Freud, biologist of the mind* (ed. F.J. Sulloway), p. 304. Basic Books, New York.
14. Ellis, H. (1936). *Studies in the psychology of sex*. New York.
15. Kinsey, A.G., Pomeroy, W.B., and Martin, C.E. (1948). *Sexual behavior in the human male*. Saunders, Philadelphia.
16. Darwin, C.R. (1859). *On the origin of species by means of natural selection, or, the preservation of favored races in the struggle for life*. John Murray, London.
17. Lashley, K.S. (1938). The experimental analysis of instinctive behavior. *Psychology Review*, **45**, 445.
18. Money, J. and Lamacz, M. (1989). *Vandalized lovemaps: paraphilic outcome of seven cases in pediatric sexology*. Prometheus Books, Amherst, NY.
19. Fedoroff, J.P., Fischell, A., and Fedoroff, B. (1999). A case series of women evaluated for paraphilic sexual disorders. *The Canadian Journal of Human Sexuality*, **8**, 127–40.
20. Blanchard, R. and Bogaert, A.F. (1996). Homosexuality in men and number of older brothers. *The American Journal of Psychiatry*, **153**, 27–31.
21. Gaffney, G.R., Lurie, S.F., and Berlin, F.S. (1984). Is there familial transmission of pedophilia? *The Journal of Nervous and Mental Disease*, **172**, 546–8.
22. Fedoroff, J.P. and Pinkus, S. (1996). The genesis of pedophilia: testing the abuse to abuser hypothesis. *Journal of Offender Rehabilitation*, **23**, 85–101.
23. Shapiro, C.M., Trajonovic, N.N., and Fedoroff, J.P. (2003). Sexomnia-a new paraphilia? *Canadian Journal of Psychiatry*, **48**, 311–7.
24. Bradford, J. (1987). Medical interventions in sexual deviance. In *Sexual deviance* (eds. D.R. Laws and W. O'Donohue), pp. 449–64. The Guilford Press, New York.
25. Grossman, L.S., Martis, B., and Fichtner, C.G. (1999). Are sex offenders treatable? A research overview. *Psychiatric Services*, **50**, 349–61.
26. Bradford, J.M.W. (2001). The neurobiology, neuropharmacology and pharmacologic treatment of the paraphilias and compulsive sexual behaviour. *Canadian Journal of Psychiatry*, **46**, 26–34.
27. Greenberg, D.M and Bradford, J.M.W. (1997). Treatment of the paraphilic disorders: a review of the role of selective serotonin reuptake inhibitors. *Sexual Abuse: A Journal of Research and Treatment*, **9**, 349–60.
28. Fedoroff, J.P. (1994). Serotonergic drug treatment of deviant sexual interests. *Annals of Sex Research*, **6**, 105–17.
29. Fedoroff, J.P., Peyser, C., Franz, M.L., *et al.* (1994). Sexual disorders in Huntington's disease. *The Journal of Neuropsychiatry and Clinical Neurosciences*, **6**, 147–53.
30. Marshall, W.L., Marshall, L.E., Serran, G.A., *et al.* (2006). *Treating sexual offenders*. Routledge, New York.

31. Hanson, R.K., Gordon, A., Harris, A.J.R., *et al.* (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment of sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, **27**, 169–95.
32. Marques, J.K., Weideranders, M., Day, *et al.* (2005). Effects of a relapse prevention program on sexual recidivism: final results from California's Sex Offender Treatment and Evaluation Progject (SOTEP). *Sexual Abuse: A Journal of Research and Treatment*, **17**, 79–107.
33. Laws, D.R. and O'Donohue, W. (eds.) (1997). *Sexual deviance: theory, assessment, and treatment*. The Guilford Press, New York.
34. Fedoroff, J.P., Kuban, M., and Bradford, J. (in press). Laboratory measurement of penile response in the assessment of sexual interests. In *Sex offenders: a multi-disciplinary approach to identification, risk assessment, treatment and legal issues* (eds. F. Saleh, J. Bradford, A. Grudzinskas, and D. Brodsky). Oxford Press, New York.
35. Grubin, D. (in press). Using the polygraph to manage risk in sex offenders. In *Assessment and treatment of sexual offenders: a handbook* (eds. A.R. Beech, L.A. Craig, and K.D. Browne). Wiley, London.

4.11.4 Gender identity disorder in adults

Richard Green

History

The behavioural phenomenon of transsexualism (now gender identity disorder) is ancient. It has been recorded for centuries and in a broad range of cultures.[1] The historic behavioural picture is comparable to that seen clinically.

In the first half of the twentieth century medical reports of sex reassignment surgery were described in Europe, primarily in Switzerland.[2] In the 1930s a wide-selling biography *Man into Woman* described a Dutch painter who underwent surgical sex reassignment.[3] Contemporary interest in transsexualism surged in 1952 when George Jorgensen, an American, travelled to Denmark and underwent hormonal and surgical treatment to become Christine Jorgensen.[4] The resultant international publicity yielded hundreds of people worldwide applying to the Danish doctors for similar treatment.[5]

By the mid-1960s there were surgeons scattered in several countries performing sex reassignment. Then in the United States at the Johns Hopkins Hospital and the University of Minnesota Hospitals and in the United Kingdom at Charing Cross Hospital, comprehensive sex reassignment programmes commenced. Extensive publicity was given to the Johns Hopkins programme as initially reported in the *New York Times* in 1966. It described the rationale for the programme and in the words of its director, 'if the mind cannot be changed to fit the body, then perhaps we should consider changing the body to fit the mind'.[6]

In 1966 the first professional text on transsexualism was written by Harry Benjamin, widely acknowledged as the 'father of transsexualism'.[7] In 1969 the first multidisciplinary text was edited by the author and John Money.[6] During the past 35 years the recognition of transsexualism, or gender identity disorder, as a treatable condition requiring psychiatric, endocrine, and surgical intervention has been accepted.

Epidemiology

The prevalence of gender identity disorder in adults is estimated from a comprehensive appraisal in the Netherlands at 1 in 10 000 males and 1 in 30 000 females.[8] At nearly all clinical centres the ratio of male-to-female patients ranges from 3:1 to 4:1, in favour of males. In some East European centres the ratio is 1:1 or reversed.[9]

Diagnosis

Diagnostic criteria of gender identity disorder in adults in DSM-IVTR[10] include a stated desire to be the other sex, a desire to live and be treated as the other sex, or the conviction that he or she has the typical feelings and reactions of the other sex. There is a preoccupation for removal of primary and secondary sexual characteristics and for procedures to alter physically the sexual characteristics to simulate the other sex. The condition is not associated with physical intersex. ICD-10 diagnostic criteria are similar but there is no mention of intersex exclusion.[11]

Origins

The search for the origins of transsexualism continues with an increasing bias towards those that are physiological. Some 20 years ago there was a false prophet in the guise of the HY antigen, on the Y chromosome believed to be influential in the development of the testes. A series of male transsexuals were found to be lacking this antigen and the tentative conclusion reached was that its absence resulted in a failure to masculinize the brain in the direction of a male identity.[12] However, the author's collaborative effort to replicate that study was not successful as all the male transsexuals studied appeared to have normal HY antigen.[13]

A more recent finding from the Netherlands implicates the brain region known as the bed nucleus of the stria terminalis. In a series of six male transsexuals studied at post-mortem over a 10-year period the size of the nucleus was comparable to that of typical females and not males.[14] A criticism of this study is that the long-term oestrogen treatment for these males may have altered the size of the nucleus. In response the researchers argue that males treated with anti-male hormone drugs or oestrogen for prostate cancer do not have an alteration in the nucleus size from typical males. However, this treatment may not be comparable to that given to transsexuals. Another criticism is that the sex difference in size of the nucleus does not manifest until early adulthood whereas the symptoms of GID often manifest earlier.

Research with male transsexuals has revealed what might be indirect markers reflecting biological distinctions. In agreement with other researchers' findings that male homosexuals have a greater likelihood of having older brothers,[15] our homosexually oriented male transsexuals also have more older brothers.[16] A theory behind this finding is that there is a progressive immunization with each pregnancy by the pregnant mother against the male foetus reflecting antigenicity of the Y chromosome. This would disrupt typical male development.

A higher ratio of aunts to uncles on the mother's side has also been found in our male transsexuals[17] a finding previously

reported by another researcher for male homosexuals.[18] A theory here is that a semilethal factor has been operant in one generation (against uncles) that in the subsequent generation influences brain development resulting in an atypical behavioural pattern (homosexual or transsexual development). The finding is explainable with genomic imprinting where a gene can be dormant in one generation depending on which parent transmitted it.[17]

We also find that both male and female transsexuals are more often non-right-handed.[19] Hand use preference begins in utero and may reflect hormonal levels or cerebral dysfunction.

For female transsexuals, a series of reports indicates a higher rate of polycystic ovarian disease.[20] Although such women secrete higher levels of androgen than typical females in adulthood, prenatal levels are unstudied. However, nearly all patients with polycystic ovarian disease are not transsexual and most female transsexuals do not have polycystic ovarian disease.

Treatment

There have been no randomized controlled trials of treatment and clinical management has evolved from decades of experience.

Prior to recognition of transsexualism as a disorder deserving medical and psychiatric attention many patients self-mutilated or committed suicide[7] Transsexual patients are helped by sympathetic assessment and intervention. However, transsexuals can be difficult patients to treat. It is a rare disorder in which patients make their own diagnosis, 'I am transsexual', and prescribe their own treatment, 'I want sex-change hormones and surgery'. Patients can be demanding and impatient for the therapist's acquiescence. They may be resentful for having to see a psychiatrist, holding the opinion that the desire for hormonal treatment and surgery should be sufficient, and psychiatric agreement should be unnecessary. Some patients will threaten self-mutilation or suicide if their demands and time schedules for demands are not met. Patients need to know that psychological stability is a key ingredient to successful negotiation of the cross-gender living trial period 'Real Life Experience' (see below) and recommendation for surgery, and that suicidal behaviour is a contraindication to going forward.

The general psychiatrist should take a full psychosexual history with emphasis on the onset and development of the gender dysphoria, attempts at treatment, and the patient's long-term goals and appreciation of the obstacles to be confronted. General psychiatric and medical status needs to be understood. A referral is to be made to a specialist centre.

During the past 30 years medical doctors and psychologists specializing in the treatment of transsexualism have worked to develop effective intervention strategies. Early on there was some optimism that extensive, prolonged psychotherapy could modify the patient's gender identity to conform to the patient's birth sex. However, in the vast majority of cases, this was not possible.

During the past 20 years one project undertaken by the Harry Benjamin International Gender Dysphoria Association, the professional body dedicated to the study and treatment of transsexualism, has been setting an extensive series of requirements for evaluation and treatment of gender identity disordered people. This set of guidelines is known as the *Standards of Care*.[21] Their principal purpose is to assure that people presenting with dissatisfaction continuing to live in the sex role to which they were born undergo comprehensive psychiatric and other medical evaluation and enter into an appropriate treatment programme. The programme includes, in addition to ongoing psychiatric or psychological monitoring, possibly endocrine therapy and, depending on the outcome of the graduated trial period of cross-gender living, possibly sex reassignment surgical procedures. The philosophy of treatment is to do reversible procedures before those that are irreversible. Thus, clothing change, name change, and cross-gender role socialization, would precede endocrine treatment with its gradual somatic changes, followed, in carefully selected cases, by surgical treatment.

The screening and evaluation of patients given the opportunity to demonstrate that they will benefit from cross-gender living, perhaps culminating in surgical treatment, is known as the 'Real Life Experience'. This requires that patients live full-time for at least a year and preferably 2 years in that role. The experience includes high doses of cross-sex hormones and full-time employment or full-time student status in the new role for at least a year. If patients can demonstrate to themselves and mental health experts that they have successfully negotiated the 'Real Life Experience' and are adjusting better in this new gender role, they can be referred for surgery.

Hormonal effects

Response to hormonal treatment is variable. This is particularly notable and potentially problematic for males. As with people born female, breast development spans a continuum. Patients may erroneously believe that more oestrogen will result in greater breast development. They neglect the fact that people born female have quite adequate female hormone production but the limiting factor is tissue response. In addition to breast development, male patients report increased hip and buttock fat, skin softening, and loss of sex drive and erection capacity. Vocal retraining is required and perhaps surgical alteration of the larynx to effect a woman's voice. Facial hair removal is required. Some patients opt for major facial contour reconfiguration, performed by a few cosmetic surgeons.

Androgen treatment to the female results in voice deepening, facial hair growth, general body hair growth, menses cessation, clitoral hypertrophy, and increased sex drive. Testosterone effects are very pronounced. The deepening voice and beard growth and perhaps scalp hair loss can metamorphose the female's appearance dramatically.

Surgery

Genital surgical treatment for the male includes penectomy, orchidectomy, and creation of a neovagina. The neovagina may be created from penile skin, perhaps augmented by other skin, or from a portion of large intestine.[22] The cosmetic result is usually very good. The extent of sexual responsivity with the neovagina is mainly anecdotal. Many patients report the subjective experience of orgasm but describe it in a different form from that experienced prior to surgery as a male. Very few physiological measures of the sexual response cycle have been reported with postoperative transsexual patients.[23]

Female transsexuals undergo bilateral mastectomy, and usually hysterectomy and ovariectomy. Genital surgery is an option taken

by perhaps half because of the limitations of phalloplasty. Two major approaches are utilized. One is creation of a micropenis from the androgen enlarged clitoris with relocation of the urethra to enable micturition while standing. Prosthetic testes can be incorporated into the labia sutured together. The microphallus will not permit vaginal penetration for intercourse but is erotically sensitive.[24] Alternatively, phalloplasty involves major surgical interventions with scarring at donor sites, particularly the arm.[25] The neophallus is not as close cosmetically to a natural penis as some patients want. It can be made more rigid with an inflatable implant and a conduit for urine may be surgically created. A procedure anastomosing a nerve from the arm to that enervating the clitoris offers promise of erotic sensation along some of the neophallus.

Sex reassignment outcome

Follow-up reports on operated transsexuals are generally quite favourable. An early review of several follow-up studies,[26] reported on 283 male-to-female transsexuals. Results were judged satisfactory for 71 per cent, uncertain for 17 per cent, and unsatisfactory for 12 per cent. For 83 female-to-male transsexuals results were judged satisfactory for 81 per cent, uncertain for 13 per cent, and unsatisfactory for 6 per cent. A more recent report considered reassignment successful in 46 of 50 male-to-female transsexuals and successful in all 61 female-to-male transsexuals.[27] In another study, of 55 male-to-female transsexuals, none regretted surgery and none had significant doubts regarding their reassignment status as women. Of 25 female-to-male transsexuals, at least 90 per cent were judged successful.[28] In a review of the English language literature over a 10-year period for operated transsexuals, 90 per cent of male-to-female transsexuals were judged to be satisfactory and 95 per cent of female-to-male transsexuals were similarly judged successful.[29] The criteria for success in some studies include objective measures of psychological and vocational status, in others the criterion is limited to an absence of regret over the reassignment process.

One study is notable because it effected some randomization of treatment conditions.[30] It was reported on 40 male-to-female transsexuals approved for surgery at Charing Cross Hospital, London. As patients qualified for surgery they were randomly assigned to two groups. Half were operated on in 3 months and the other half were kept on a waiting list for about 2 years. All patients completed a standardized assessment at acceptance for surgery and at the end of 2 years. The group that received the earlier surgery showed significant improvement in a range of psychometric measures and maintained employment. The unoperated group showed no improvement in psychological testing and deteriorated in employment.

Family management

Transsexual patients are often married and have children. When possible, the family should become part of the treatment process. Typically relations are strained with the marital partner of the transsexual and divorce is usual. Particularly when children are involved an effort should be made to deal with the feelings of betrayal or abandonment from the transsexual's spouse that contaminate the continuing relationship between transsexual parent and children. Often there is concern by the parents that the patient's transsexualism will impact adversely on the children. There is concern specifically in areas of gender identity of the children and peer group reactions to the knowledge that one of their age mate's parents is transsexual. However, in the author's research of 34 children of transsexuals who were living with or in regular contact with the transsexual parent, there were no instances of gender identity disorder in the children and no instances of peer group alienation that were especially problematic.[31] Children typically have many questions about the transsexual transformation of that parent that can be answered by the clinician, perhaps with the transsexual present.

The third sex

A recent development in the pattern of patients presenting clinically are those with a transgendered identity, popularly known as 'the third sex'. These males or females do not request 'sex change'. Rather, they want, if male, to be demasculinized and, if female, to be defeminized. Thus males may want castration and penectomy but no oestrogen treatment and no vagina, and females may want mastectomy, perhaps hysterectomy, but no androgen treatment and no phallus.

These patients pose a dilemma for clinicians. The crux of patient management for gender identity disorder is the 'Real Life Experience' (see above), including cross-sex hormonal treatment, the prelude to possible surgical alteration. Reversible procedures precede those that are irreversible in this management strategy. But with third-sex patients, no 'Real Life Experience' is possible. They do not have a trial period. Guidelines for testing the rationality and stability of their requests need to evolve from the body of clinicians currently attempting management of this unique population.

Transsexual patient subgroups

Professionals not experienced in the treatment of large numbers of male-to-female transsexual patients are often unaware that a substantial minority, perhaps a third, are not sexually oriented to male partners. Many of these patients have been married and are fathers, and many will be bisexual or remain sexually attracted to females only after surgery and live as lesbian women. A much smaller number of female-to-male transsexuals are sexually attracted to male partners and live as gay men after reassignment surgery.

In addition to the subtypes of transsexuals based on their sexual orientation, some male transsexual patients evolve through a diagnostic phase more closely fitting fetishistic transvestism. These patients have been more masculine in general lifestyle and appearance than other male transsexuals, cross-dressing has been sexually arousing, and they have usually been heterosexually oriented. However, with the passage of time gender dysphoria increases and fetishistic components of cross-dressing diminish or disappear. Many have been sexually aroused by fantasies of themselves as women.[32] They have been termed 'autogynephiles'. There is some evidence that males evolving through a fetishistic cross-dressing phase, presenting as somewhat older at gender identity clinics, have a poorer prognosis after surgery. However, it is primarily the progression through the 'Real Life Experience' that becomes the critical management guideline for patients, irrespective of their background.

Gender identity as a disorder

A growing movement among transsexual people argues for removal of gender identity disorder from the psychiatric and medical lists of disorders or diseases. These advocates argue that their sexual identity is normal male or female and that surgical correction of their anatomical anomaly is all that is required to allow them to live as normal men or women. Their condition is depicted as distinctive from the recognized traditional forms of mental disorder such as schizophrenia or major depression. Furthermore, carrying a psychiatric diagnosis is stigmatizing. One argument for inclusion of transsexualism in the APA diagnostic manual of disorders is that it follows the criteria of other entries, that is subjective distress and social disadvantage. Another considers third-party payment for treatment. It is unlikely that medical insurance carriers, private or governmental, would fund intervention for a non-medical condition.

Additionally, there is objection by some persons with gender identity disorder to being referred to as transsexual. They prefer the designation 'transwoman' for what has been known as male-to-female transsexualism, and 'transman' for female-to-male transsexualism.

Legal issues

Cultural and legal approaches to transsexualism vary widely across nations and cultures. They are beyond the scope of this chapter. However legal issues in the United States and United Kingdom can be summarized as follows.

Part of the 'Real Life Experience' of cross-gender living includes employment in the desired gender role. However, transsexuals may be the object of employment discrimination based on their transsexualism. In the United States, transsexuals were denied federal protection 20 years ago when the court held that the anti-sex discrimination statute protected men and women, but not transsexuals.[33] Finally, in 2004, another federal court extended employment protection.[34] In the United Kingdom, in 1997, a ruling on an English case before the European Court of Justice held that anti-sex discrimination law, to which all European Union members were subject, did include transsexuals. Thus discrimination in employment was illegal.[35]

Changing sex on one's birth certificate can be an important step in the life of the postoperative transsexual. In its absence, the person's legal sex may remain in the preoperative status, and pose obstacles to a full life. Marriage is a key issue. In the United Kingdom, until recently, postoperative transsexuals could not have a new birth certificate issued and change their legal sex. Thus a male-to-female transsexual could not marry a male and a female-to-male transsexual could not marry a female. This changed in a statute enacted in 2004.[36] In the United States most states permit some birth certificate change. However, that change may not be recognized in another state.

Further information

Barrett, J. (ed.) (2008). *Transsexual and other disorders of gender identity*. Radcliffe Publishing, Oxford.

Ettner, R., Monstrey, S., and Eyler, E. (eds.) (2008). *Principles of transgendered medicine and surgery*. Haworth Press, Binghampton.

References

1. Green, R. (1966). Transsexualism: mythological, historical and cross-cultural aspects. In *Appendix to The transsexual phenomenon* (ed. H. Benjamin). Julian Press, New York.
2. Abraham, F. (1931). Genital alteration in two male transvestites. *Zeitschrift Sexualwissenschaft*, **18**, 223–6.
3. Hoyer, N. (1933). *Man into woman: an authentic record of a change of sex*. E.P. Dutton, New York.
4. Hamburger, C., Sturup, G., and Dahl-Iversen, E. (1953). Transvestism: hormonal, psychiatric and surgical treatment. *The Journal of the American Medical Association*, **12**, 391–6.
5. Hamburger, C. (1953). The desire for change of sex as shown by personal letters from 465 men and women. *Acta Endocrinologica*, **14**, 361–75.
6. Green, R. and Money, J. (eds.) (1969). *Transsexualism and sex reassignment*. Johns Hopkins Press, Baltimore, MD.
7. Benjamin, H. (1966). *The transsexual phenomenon*. Julian Press, New York.
8. Kesteren, P., Gooren, L., and Megens, J. (1996). An epidemiological and demographic study of transsexuals in The Netherlands. *Archives of Sexual Behavior*, **25**, 589–600.
9. Godlewski, J. (1988). Transsexualism and anatomic sex ratio reversal in Poland. *Archives of Sexual Behavior*, **17**, 547–8.
10. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, text revision). American Psychiatric Association, Washington, DC.
11. World Health Organization. (1992). *International statistical classification of diseases and related health problems, 10th revision*. WHO, Geneva.
12. Eicher, W., Spoljar, M., Cleve, H., *et al.* (1979). H-Y antigen in transsexuality. *Lancet*, **2**, 1137–8.
13. Wachtel, S., Green, R., Simon, N., *et al.* (1986). On the expression of H-Y antigen in transsexuals. *Archives of Sexual Behavior*, **15**, 49–66.
14. Zhou, J., Hoffman, M., Gooren, L., *et al.* (1995). A sex difference in the human brain and its relation to transsexuality. *Nature*, **378**, 68–70.
15. Blanchard, R. (1997). Birth order and sibling sex ratio in homosexual versus heterosexual males and females. *Annual Review of Sex Research*, **8**, 27–67.
16. Green, R. (2000). Birth order and ratio of brothers to sisters in transsexuals. *Psychological Medicine*, **30**, 789–95.
17. Green, R. and Keverne, E.B. (2000). The disparate maternal aunt–uncle ratio in male transsexuals: an explanation invoking genomic imprinting. *Journal of Theoretical Biology*, **202**, 55–63.
18. Turner, W. (1995). Homosexuality, type 1: an Xq 28 phenomenon. *Archives of Sexual Behavior*, **24**, 109–34.
19. Green, R. and Young, R. (2001). Hand preference, sexual preference, and transsexualism. *Archives of Sexual Behavior*, **30**, 565–74.
20. Futterweit, W., Weiss, R., and Fagerstrom, R. (1986). Endocrine evaluation of 40 female-to-male transsexuals: increased frequency of polycystic ovarian disease in female transsexualism. *Archives of Sexual Behavior*, **15**, 69–78.
21. Harry Benjamin International Gender Dysphoria Association. (1981). *The standards of care for gender identity disorder*. Symposion, Düsseldorf.
22. Schrang, E. (1998). Male-to-female feminizing genital surgery. In *Current concepts in transgender identity* (ed. D. Denny), pp. 315–33. Garland, New York.
23. Green, R. (1998). Sexual functioning in post-operative transsexuals. *International Journal of Impotence Research*, **10**(Suppl. 1), 22–4.
24. Hage, J. and van Turnhout, A. (2006). Long-term outcome of metaidoiplasty in 70 female-to-male transsexuals. *Annals of Plastic Surgery*, **57**, 312–16.

25. Monstrey, S., Hoebeke, P., Dhont, M., *et al.* (2005). Radial forearm phalloplasty, review of 81 cases. *European Journal of Plastic Surgery*, **28**, 206–12.
26. Pauly, I. (1981). Outcome of sex reassignment surgery for transsexuals. *Australian and New Zealand Journal of Psychiatry*, **15**, 45–51.
27. Blanchard, R., Steiner, B., Clemensen, L., *et al.* (1989). Prediction of regrets in postoperative transsexuals. *Canadian Journal of Psychiatry*, **34**, 43–5.
28. Kuiper, B. and Cohen-Kettenis, P. (1988). Sex reassignment surgery. A study of 141 Dutch transsexuals. *Archives of Sexual Behavior*, **17**, 439–57.
29. Green, R. and Fleming, D. (eds.) (1991). Transsexual surgery follow up: status in the 1990s. In *Annual review of sex research* (eds. J. Bancroft, C. Davis, and D. Weinstein). Society for the Scientific Study of Sex, Mount Vernon, IA.
30. Mate-Cole, C., Freschi, M., and Robin, A. (1990). A controlled study of psychological and social change after surgical gender reassignment in selected male transsexuals. *The British Journal of Psychiatry*, **157**, 261–4.
31. Green, R. (1998). Transsexuals' children. *International Journal of Transgenderism*, **2**, 1–7.
32. Blanchard, R. (1989). The classification and labelling of non-homosexual gender dysphorics. *Archives of Sexual Behavior*, **18**, 315–34.
33. Green, R. (1986). Spelling relief for transsexuals: employment discrimination and the criteria of sex. *Yale Law and Policy Review*, **4**, 125–40.
34. Smith,v. City of Salem. (2004). 378 F.3d 566 (6th Cir.)
35. P v. S and Cornwall County Council. (1996). ECRI-2143 C-13/94.
36. Gender Recognition Act. (2004).

4.12

Personality disorders

Contents

4.12.1 Personality disorders: an introductory perspective

Juan J. López-Ibor Jr.

The goal of psychiatry is the study of mental illnesses. In this chapter we consider the degree to which personality disorders can be considered as mental illnesses.

Basic notions

Personality is the quality that makes each one of us both different from others and consistently recognisable throughout our lives. Hence, there are two approaches to study personality. One is transversal, consisting on description of archetypes of human beings. One of the first to take this approach was Theophrastus (372–287/5 BC) who in his book *The Characters*, portrays thirty-two such prototypes. Some of them can be are easily recognized by present-day psychiatrists, for instance those typified by poor impulse control: The offensive man (*bdeluria*), the unsociable man (*authadeia*), the show-off (*alazoneia*) and the slanderer (*kakologia*); by obsessive traits: the superstitious man (*deisidaimonia*) or by paranoid traits: the suspicious man (*apistia*). The corresponding contemporary approach consists of the isolation of psychological traits or dispositions, to describe permanent inclinations to behave in a preset way.

The longitudinal approach to the study of the personality is based on the notion that there is an initial seed that develops through the lifetime. Sir Francis Galton (1822–1911) was among the first to consider the inheritance of individual differences in humans, although for centuries breeders of dogs, horses or bulls for bullfighting, had been selecting animals for mating on to select desired characteristics whether is be hunting, running or fighting.

Twin and developmental studies have been used. For example, The 'New York Longitudinal Study'[1] on infant temperament started in the early 1950s and examined how temperament influences adjustment throughout life. Kagan *et al.*[2] followed up a cohort of babies to age 14-17 years and reported that those who were highly reactive when they were babies were more likely to be 'subdued in unfamiliar situations, to report a sour mood and anxiety over the future and to be more religious'.

There are two key features of personality, one of which is temperament and the other character. The two together constitute personality.

Temperament is the innate predisposition to behave in a particular manner. Historically the concept was part of the theory of the four humours, which had corresponding temperaments: *sanguine* (the individual is led by his own pleasure to live), *choleric* (the individual has a feeling of power and shows it), *melancholic* (the individual is dominated by doubts and ruminations) and *phlegmatic* (the individual lacks any links to life, lives without effort nor pleasure). Current research on the biological basis of personality has renewed the interest in temperament.

Character is a configuration of habits, a disposition, consisting in the actualised aspects acquired through learning and shaped by experience.

Psychiatry and abnormal behaviours

Descriptions of individuals with behavioural characteristics of a negative moral or social value exist in every culture and most societies have established institutions in which the marginalized have been confined, as recorded by Foucault.[3] The distinction between immoral behaviour and mental illness was established in France at the end of the eighteenth century, coinciding with the birth of modern psychiatry. The Marquis de Sade was expelled from the Chârenton Hospital because, in words of the director, 'he is not ill, his only madness is vice'.[3] Pinel (1745–1826) considered that, in the case of the young man who in an attack of rage threw a woman into a well, although his ability to judge was clear and intact and although he presented no delusional ideas, his behaviour was characteristic of a mental patient. Consequently, this murderer was diagnosed as suffering from *manie sans délire* and his madness was classified as **reasoning madness** (*folie raisonnante*).[4] This reasoning is similar to that of Cleckley 150 years later who proposed that the social maladaptation of psychopaths is of such high degree that should considered as the result of an underlying psychotic disturbance, being personality disorders are a *mask of sanity*.[5]

Prichard,[6] defined the concept of moral insanity from which, together with the moral degeneration described by Morel (1809–1873),[7] the modern concepts of psychopathy and personality disorders are derived.

Difficulties in the study of personality disorders

Two factors have prevented the development of scientific knowledge in this field: first the negative evaluation of the concept of moral insanity, and second the dualism inherent in psychopathology.

The stigma of personality disorder

The diagnosis of personality disorder generally implies the idea of intractability and frequently leads to a lack of proper medical care. This attitude is the expression of a negative, moralising, and, according to Tyrer *et al.*,[8] delusional attitude of the doctor towards the patient. Cusack and Malaney[9] posed the question as to whether patients with antisocial personality disorders are 'bad' or 'mad'. They attempted to establish differential criteria in order to show that if patients with an antisocial personality disorder are not 'mad', then they must be considered as 'bad' and therefore must be delivered to the judicial system, after diagnosis and treatment of secondary symptoms. In 1999, the UK Government 1999 introduced a new concept: Dangerous and Severe Personality Disorder (DSPD). This subsequently became a treatment and assessment program for individuals who satisfy three requirements: 1) have a severe disorder of personality, 2) present a significant risk of causing serious physical or psychological harm from which the victim would find it difficult or impossible to recover, and 3) the risk of offending should be functionally linked to the personality disorder.[10, 11]

Dualism in psychopathology

Dualism has been present in psychiatry since its origins as specialty. According to Griesinger:[12]

> *It is time that [mental medicine] should be cultivated as a branch of brain pathology and of [the study of] the nervous system in general, and to apply serious diagnostic methods used in all branches of medicine. . . . Besides this purely medical element, mental medicine has another essential one and which gives a special and proper character to this part of the healing art; it is the psychological study of the aberrations of the intelligence observed in mental illnesses.*

The radical separation between mental-brain illnesses and 'aberrations of intelligence' is fundamental to modern psychopathology. Schneider[13] distinguished between psychoses as pathological conditions of the brain (disease or defective structure) and variations of the psychic way of being. Abnormal personalities, personality disorders, and neurotic disorders belong to the second category.

Schneider[14] defines some **abnormal personalities** in a statistical sense, to describe those individuals whose form of feeling, experience and behaving differs to a certain degree from what is considered to be normal for most individuals in a social group. Some of these are **psychopathic personalities** who, as a result of their abnormality, suffer or make others suffer. It should be stressed that according to Schneider's statistical definition of personality and the dualism of his psychopathological system, the only possible criterion to define a clinical condition in the absence of a brain disease is the suffering, the pathos. Suffering is for Schneider the reason why some people ask for medical care, but not a sufficient criterion for to determine the presence of an illness. Schneider had to add suffering inflicted on others (social suffering) in order to be able to include certain kinds of abnormal personalities characterized by the absence of personal suffering (heartless psychopaths, sociopaths).

It seems acceptable to consider as a patient someone who suffers and asks for clinical care although the criterion for suffering is a weak one when compared with the presence of an illness of an organ. On the contrary, the criterion of induced suffering which characterises some psychopathic personalities, defined following Schneider, is not acceptable in medicine and it is surprising that this has been little criticised. The clue lies in Schneider's definition of personality which *excludes* any biological substrate.[15] The effect of viewing psychopathies as simple variations was to reduce the amount of neurobiological research into the neuroses and personality disorders because they were not considered amenable to natural scientific methods. The study of the personality was left the new psychoanalytical and psychological theories.

Nowadays it is impossible to maintain such a reductionistic perspective, and it is recognised that the morbid nature of personality disorders can be understood through the study of changes in its biological substratum. There are not two kinds of mental disorders, the psychosis which are the consequences of brain illnesses, and the variations of the psychological way of being (neurosis and personality disorders), but two inherent aspects to each disorder. It is essential to consider psychological and psychopathological aspects of psychoses, as well as the brain dysfunction of the variations of the psychic way of being.

Models of personality

The study of personality by the different schools of differential psychology provides a solid background to help to understand the disorders of personality. Unfortunately, these studies have been

conducted from different and sometimes contradictory perspectives, which are summarized in the following sections.

Categorical perspective

The categorical perspective is deep rooted in the psychiatric tradition. Categorical models consider discontinuous personality categories. This type of model is used in DSM-IV[16] and ICD-10[17] because of the need for a specific diagnostic, i.e. a categorical approach.

In modern nosology the categorisation of illness is based on the symptoms present and not on their aetiopathology, and says nothing about the nature of the disorders themselves. In the case of personality disorders, the categorisation does not affirm or deny that they are disorders or illnesses, nor does it indicate where the symptoms differ from non-morbid behaviour patterns.

This approach is supported by the notion of ideal types of Weber (1864- 1920), introduced into psychiatry by Jaspers (1883–1969) and more recently by Schwartz and Wiggins.[18] Ideal types are constructs to understand reality: An ideal type is formed by a unilateral accentuation of one or more perspectives and by the synthesis of a great deal of individual phenomena. A type describes the perfect case. Recently Doerr[19] has argued that the ideal types, when well described, become almost real types.

The experimental approach

The experimental approach looks for general laws on personality and establishes causal relations between personality variables. Wundt (1832–1920) studied the effects of modifications of stimuli on the intensity and quality of the subject's experiences introduced them. Pavlov (1849–1936) studied the conditioning of the responses to stimuli and the experimental neurosis. The behavioural approach to the personality was introduced by Watson (1878–1958) who applied objective methods to the study of human behaviour and to the relationship between stimuli and responses. Hull (1884–1952) expanded behaviourism to include learning, feelings, expectations, achievements, goals and motivations. This led to the notion that the stimulus response relationship is influenced by cognitive processes. Perception, memory, language and other functions influence the processing of information of the surrounding world and the information coming from oneself (*self*). Skinner (1904-1990) created a theory of the operating conditioning, result of a non-adaptative learning. From these perspectives, the personality is viewed as a computer which introduces, stores, transforms and produces information, including the contents of the information as well as the process in itself.[20]

The psychoanalytical approach

Freud (1856-1939)[21] proposed in the course of his life three different models of personality: The first was the speculative neuropsychological model of the *Project of a psychology for neuropsychiatrists* (1897) based on the concepts of psychic energy and psychodynamic. The second was the topographic model of the *The Interpretation of Dreams* (1900) where Freud described the conscious, preconscious and unconscious levels. The third is the structural model of *The Ego and the Id* (1923) and of *Inhibition, Symptoms and Anxiety* (1926) where Freud introduced the notions of the Id, the Ego and the Super-Ego. This last model led to a new perspective, the psychology of the Ego and the description of defence mechanisms which have a strong impact on the study of personality in clinical settings. Defence mechanisms distort reality to adapt the subject to it and to reduce anxiety. Some of them are more normal (promote adaptation to the environment), others are pathological (maladjusted or maladaptative).

The correlational approach

The correlation approach explains the individual differences based on personality traits and applying a dimensional model based on statistical correlation. Karl Pearson (1857–1936), the founder of mathematical statistics introduced the correlation coefficient (correlates of cognitive flair with variables like age, gender, weight, height and so on). Charles E. Spearman (1863–1945) applied to research on the traits of personality, a factorial analysis that groups different qualities around a series of correlational factors or dimensions.

Traits are the basic elements of a personality and individual differences are defined and classified along dimensions. The theoretical assumption is that the structure of personality is common to all individuals; it differs in the different combination of traits. Trait is a disposition to respond in a determined way to a determined situation. Traits characterize persons through brief and precise descriptions on stable ways to behave, and as behaviour is consistent, it is possible to predict behaviours. This approach has paved the way to basic dimensions of individual differences.

Dimensional models

Jung (1875-1961) made the first important contribution to the dimensional concept of the personality, based on the concept of trait or disposition.[22] A trait is a permanent inclination towards behaving in a determinant way. Traits are distributed along dimensions which make it possible to classify individuals according to their personality. The different dimensional models are based on the supposition that we all share the same personality structure, differing in the different combination of the mentioned traits. These models have benefited from the innovative statistic techniques, which allows to group different qualities of the individual character around factors of correlation or dimensions.

This dimensional approach raises several questions. How many traits define personality? Are the traits universal? Do traits relate only to manifest behaviours or are they part of feelings, values or thoughts? The problem of the number of dimensions that define personality led to the search for external validators such as biological, cultural and genetic factors. Eysenck[23] identified there are three basic types of personality: extroversion-introversion, neuroticism and psychoticism, each one including multiple levels of traits. For Eysenck and Eysenck,[24] the concept of arousal level is essential. Every individual has an optimal activation level of specific systems of the central nervous system—the better they feel, the better they will perform. This approach has been developed by many authors including Zuckermann,[25] who described sensation-seeking behaviour, Oreland *et al.*,[26] and Siever and Davis,[27] who proposed new traits and dimensions.

Cloninger[28] initially proposed three dimensions: novelty-seeking, harm avoidance, and reward dependence. Latter, he attempted to overcome the dichotomy between dimensional and categorical models by using four temperamental dimensions (novelty-seeking, harm avoidance, reward dependence, and persistence), which are life-long and stable, and three character dimensions (self-direction,

co-operation, and self-transcendency) which are variable and susceptible to environmental influences and development.[29]

The five-factor model, based on factorial studies and individual differences[30] has been widely accepted. It comprises the personality dimensions openness, conscientiousness, extraversion, agreeableness, and neuroticism, known by the acronym OCEAN. About 40 per cent of individual personality differences can be explained in terms of heredity.[31] In the five-factor model the same proportion does not apply to each factor; openness to experience appears to have the greatest hereditary input, whereas conscientiousness appears to have the least.

Mathematical tools allow recombining the data in order to find higher order factors of the Big Five. Two of them have appeared in many studies: 1) related to the Big Five trait dimensions Agreeableness, Conscientiousness, and Emotional Stability (meta-trait alpha) and 2) the dimensions Extraversion and Intellect (meta-trait beta).[32] Other have found some extra traits to be added to the Big Five, such as honesty-humility.[33]

An interesting approach is lexicographic, which is based on the examination of relations among personality-descriptive adjectives that are indigenous to various languages. They tend to reveal a structure corresponding closely to the Five-Factor Model, with some differences in the nature of the Agreeableness and Emotionality/Neuroticism factors and also in the existence of a sixth factor, Honesty-Humility.[34] This has been found in different languages such a tagalong with some differences (a Filipino extra factor resembled a Negative Valence or Infrequency dimension).[34] A study with college students yielded seven major dimensions; many of the factors were similar to recognized lexical personality factors. Big Five Conscientiousness and Neuroticism were each strongly associated with a single proverb dimension (interpreted as Restraint and Enjoys Life, respectively). Big Five Agreeableness, Extraversion, and Intellect/Imagination were all associated with several proverb dimensions. Agreeableness was most strongly associated with proverb dimensions representing Machiavellian behaviour and strong Group Ties, and both Extraversion and Intellect showed particularly notable associations with an Achievement Striving dimension. The two remaining proverb dimensions, which represented a belief that Life is Fair and an attitude of Cynicism, could not be accounted for by the Big Five.[35]

Critics of this approach have argued that a) little progress has been made in this area, b) structural models have little direct relevance for psychopathology research, c) the principal methodological tool of structural research–factor analysis–is too subjective to yield psychologically meaningful results and 4), some clinically relevant aspects, such as alexithymia or impulsivity, do not appear in the studies on the structure of personality.

Alexithymia positively correlates with Neuroticism (N) and negatively with Extraversion (E) and Openness (O), whereas no significant relations were found with Agreeableness (A) and Conscientiousness (C).[36] Impulsivity. The term impulsive madness was used in German literature and Jaspers[37] put it in relation to nostalgia and displacement. The balance between social and individual norms is related to the origin of mental disorders. Durkheim[38] introduced the concept of **anomia** when describing a particular form of suicide in individuals who perceive that their own norms and values are no longer relevant and that their relation to the community is weak or non-existent. However, the opposite may also occur. Kraus[39] coined the term **hypernomia** for premorbid personality traits of depressive patients, which consist in an exaggerated form of adaptation to social norms. This personality type is the converse of the impulsive madness that can be characterized as **hyponomic**. We have proposed the term **dysnomic** for obsessive patients[40] who show a distorted adaptation to social norms. For example, patients with obsessions and compulsions related to cleanliness usually appear to be extremely dirty because of their fear of contamination and their unrealistic compulsions, which are based more on 'magic' control than on efficient behaviour oriented towards concrete goals.

The common link in the psychopathology of obsessive-compulsive disorders and the group of impulsive disorders experienced by the **impulsivists**[41] is poor control of the impulses, in the sense that novel interior or exterior experiences are not converted into adequate behavioural patterns. Rather, obsessives abandon actions uncompleted and impulsivists behave in a disorganised manner (acting out). In both cases, 'the irrelevant substitutes the relevant'.[42]

Today, the trend is to look for a classification of personality disorder which will be dimensional, either by selecting one of the existing models by developing a common, integrative representation including the important contributions of each of the models.[43]

Models of personality and personality disorders

The ideal goal of a single structural framework to be applied to normal and abnormal personality is not easy to reach. In a study with the Eysenck Personality Questionnaire-Revised (EPQ-R) the three clusters of personality disorders of DSM found equivocal support. Exploratory principal components analysis and confirmatory factor analysis found four broad factors of personality disorder that overlapped with normal personality traits: an asthenic factor related to neuroticism; an antisocial factor associated with psychoticism; an asocial factor linked to introversion-extraversion; and an anankastic (obsessive-compulsive) factor. In spite of this, there is growing agreement about the number and type of broad personality disorder dimensions; similar dimensions may be found in clinical and non-clinical samples, suggesting that those people with personality disorders differ quantitatively rather than qualitatively from others; and there is substantial overlap between normal and abnormal personality dimensions.[44] For Livesley *et al.*,[45] personality disorders are quantitatively extreme expressions of normal personality functioning developed around four factors: Emotional Dysregulation, Dissocial Behaviour, Inhibitedness and Compulsivity. In the same year, the same group published some different results in another study:( ) they found 16 basic dispositional traits (anxiousness, affective lability, callousness, cognitive dysregulation, compulsivity, conduct problems, insecure attachment, intimacy avoidance, narcissism, oppositionality, rejection, restricted expression, social avoidance, stimulus seeking, submissiveness, and suspiciousness) and three higher-order patterns (emotional dysregulation, dissocial behaviour, and inhibitedness).[46]

Markon *et al.*[47] have delineated an integrative hierarchical account of the structure of normal and abnormal personality. This hierarchical structure integrates many Big Trait models proposed

in the literature. Similarly, O'Connor[48] reanalyzing the published studies found high level of support for both theoretically and empirically based representations of the five-factor model approach to personality disorders. The five-factor model personality dimensions of Neuroticism, Extraversion, and Agreeableness are the most apparent in the DSM-III-R conceptualizations of the personality disorders.[49,50] The five-factor model structure is present although with some variance in the current DSM-IV cluster set.[50]

Mulder and Joyce[51] have attempted to construct a simplified system for the classification of personality disorders related to normally distribute human personality characteristics. A four-factor solution of personality disorder symptoms was obtained and they labelled these factors 'the four As': Antisocial, Asocial, Asthenic and Anankastic. The factors related to the four temperament dimensions of the Tridimensional Personality Questionnaire (TPQ), but less closely to Eysenck Personality Questionnaire (EPQ) dimensions. The four factors were similar to those identified in a number of studies using a variety of assessment methods and this lends some credibility to our findings.

The masking of sanity becomes evident in some dissocial, psychopathic and even criminal behaviours that are the expression of underlying disorders. The issue is very relevant because those can be treated. One example is attention deficit hyperactivity disorder (ADHD) which has important forensic implications. For decades there has been an interest in predicting which children will become psychopaths in order to establish primary prevention interventions. Lynam[52] described the fledgling psychopath, characterized by symptoms of hyperactivity-impulsivity-attention problems and conduct problems are at the greatest risk for becoming chronic offenders. Other authors[53,54] have strongly argued against the clinical-forensic utility of tests designed to assess juvenile psychopathy.

The problems of axis II of the diagnostic and statistical manual

There are several reasons in favour of an independent axis for personality disorders. First it is important not to forget personality in the diagnostic process, especially when using symptomatic classifications (DSM-III / IV and ICD-10). Second, personality may be a predisposing factor, or something essential for the response to treatment and for prognosis.[54] Third personality traits are egosyntonic: they include traits that the subject has accepted as an integrative part of him/herself in a progressive way (when compared to axis I disorders and non-psychiatric illnesses which are something 'that occurs').

Seen from a theoretical perspective, personality disorders may be considered as the expression of the personality's functioning, which is essential for patients, be they psychiatric or not.[45]

But there are also reasons to combine Axis I and Axis II. The better a personality disorder is known and the more biological correlations are found in it, the higher the probabilities it has to be moved to Axis I: Epileptic personality (adhesivity, gliscroidy) became long ago organic disorder of the personality, cyclothymic personality became cyclothymia and depressive personality (at least in part), dysthymia in DSM-III. Schizotypal personality is schizotypal disorder in ICD-10 and several studies emphasise the strong relation of borderline personality disorder and mood (affective) disorders.[55]

Although through most of the 20th century, from K. Schneider to DSM-IV, personality disorders and mental illnesses were studied as separate fields, there has been increasing recognition of the substantial overlap of – and comorbidity between – disorders both within and across axes and interest in the joint study of normal and abnormal personality.[56] In comorbid cases, the personality disorder could be a predisposing factor, a consequence, or an attenuated form of the mental disorder, or it could be independent of the mental disorder. The fact that the association between a mental disorder and a personality disorder is not always fortuitous has been shown by the observation that effective treatment of the former can lead to the disappearance of the latter, as has been demonstrated in the treatment of obsessive-compulsive patients with pharmacotherapy and behavioural therapy.[57]

The main difference between mental and personality disorders may be that the latter are early onset variants with a very chronic course. According to the hypothesis of a spectrum of disorders, personality disorders can often be treated by the same method as those applied to the major psychiatric disorders to which they are related. Patients with anxious or avoidant personality disorder may respond to anxiolytic medication, patients with borderline personality disorder may respond to lithium and antidepressives, patients with schizotypal personality disorder may respond to antipsychotic agents, and patients with disorders characterised by poor impulse control may respond to antidepressives with a selective serotonergic action.

The diathesis-stress model

The diathesis-stress model has been used to explain the relationship between personality and mental disorders. Supposedly broad, innate temperament dimensions, sometimes correlated with somatic characteristics such as body type (Kretschmer [1888–1964]), develop and differentiate themselves through both biologically and environmental events into a hierarchical personality trait structure. At their extremes, are risk factors (diatheses) for psychopathology, especially given adverse life experiences (stress). Rosenthal[58] defined this model as an inherent constitutional predisposition which only becomes apparent under the impact of environmental stress.

The onset early in life, the variability of expression dependent on setting, the greater association with more severe disorders and the acceptance as intrinsic components of functioning by most suffering from personality disorders support the notion that they are diatheses rather than disorders.[59]

Personality disorders have been considered as belonging to the spectrum of major psychiatric disorders. However, it must be remembered that external events, such as brain damage (organic personality disorders), or the psychological impact of a catastrophic event may also lead to personality changes. Severe psychiatric disorders may have a repercussion on the personality of the patient, and other illnesses may also have this effect. For example, chronic pain (of organic nature) can be accompanied by a profound personality change (algogenic psychosyndrome). Hypochondria or dissociative symptoms and traits may become relevant only after the patient has suffered an illness, or a problem related to diagnosis or treatment, or a problem involving the patient–physician relationship.

Recent research has focused on the impact of social conditions in the neuroendocrine regulation of the individual, especially

regarding the adaptation to stressful situations. In patients with borderline personality disorder and suicidal impulsive behaviour we have found when compared to the control group, high basal concentrations of cortisol (suggesting a high level of stress) and a very blunted response to the stimulus (suggesting a reduced capacity to respond to external stimulus).[60] A clue to the interpretation of these results lies in the work of Sapolsky,[61] who has studied the adaptation to stress of baboons in the Serengeti savannah in Africa. Males of a lower rank have consistently high concentrations of the stress hormone hydrocortisone in their blood, whereas the concentration is lower in the dominant males. However, in the dominant males hydrocortisone concentrations increase rapidly at times of stress and decrease once the situation is resolved, whereas the lower-order males, who live in a permanent state of stress, are unable to initiate more adaptive resources (increase hydrocortisone secretion) when new stressful events appear. These patterns are the consequence and not the cause of the rank (if the opposite were true, the baboons who were physiologically better able to respond to stressful situations would achieve a higher rank). During periods of revolution, members of the colony hold successively different ranks and, although there is always one dominant animal, the stability of the society is lost and with it the stress-adapted physiology of the dominant males who show prolonged increased hydrocortisone concentrations like the rest of the group. When calm is established again, the normal pattern related to the hierarchical rank of the baboons is restored regardless of what their cortisol secretion pattern was prior to the revolution.

Relational disorders

This model can be applied to the relational disorders or processes. On one side every clinician has the experience of persons whose behavioural problems happen in specific environments (i.e., the family) or in relation to specific persons (i.e., the spouse) while their behaviour is totally normal in the rest of circumstances. Another observation is the case of personality disorders that once beginning to improve after a therapeutic intervention, the whole atmosphere around him or her changes. The patient himself and their relations start to consider that the personality as not a stable and incorrigible set of traits but something that can change for the better. However, it is the influence of child psychiatrists looking for childhood antecedents of mayor psychiatric disorders of adulthood[62] and family therapists that have requested a new diagnostic category to describe these situations.

It has also been claimed that in order to be able to proceed along this line, some problems have to be addressed, although the same may also be present in diagnostic categories accepted in DSM-IV and ICD-10. The main one are: 1) the little consensus on assessment means; 2) the complexity of relational assessment, which results in a lack of well-accepted, evidence-based operational definitions; 3) insufficient empirical testing of relational issues; and 4) the resistance to labeling social difficulties as disorders.[63]

These difficulties can be surmounted if new diagnostic perspectives are implemented. One is the concept of diseases as harmful dysfunctions a view which holds that disorders are harmful failures of biologically selected functions. There are evolutionarily selected functions that depend for their performance on the nature of the interaction between individuals. These relations can fail, even when both individuals are normal, because of mismatches between normal variations. Thus, there are genuine relational dysfunctions that, when harmful, are relational disorders.[64] The Structural Analysis of Social Behavior (SASB) has operationalised interpersonal theory for the research of relational aspects of psychopathology and becoming a useful diagnostic tool.[65]

The research agenda for DSM-V recognizes that the diagnosis of relational disorders is one of the most important gaps in the current DSM-IV.[66] Specific recommendations include developing assessment modules, determining the clinical utility of relational disorders, determining the role of relational disorders in the aetiology and maintenance of individual mental disorders, and considering aspects of relational disorders that might be modulated by individual mental disorders.

Personality disorders vs. personality variants

It is necessary to establish a clear distinction between personality disorders and personality variants, and to view the former as true morbid entities. ICD-10 allows us to differentiate, at least theoretically, personality disorders that appear in the chapter on mental disorders from personalities relevant for medicine in Chapter Z.[67] Personality disorders should be characterised by the presence of symptoms, and relevant personalities by their traits. Symptoms are used as diagnostic criteria in a categorical classification, while psychological traits can be classified according to dimensions.

Personality variants relevant to medicine, although not morbid, play an important role in the aetiopathogenesis of illnesses or are important for prognosis and rehabilitation. The study of the variants of personality also reminds the practitioner of the necessity to identify the uniqueness of the patient's personality.

The frontiers with normal personality are not well explained or justifiable threshold,[68] except for schizotypal disorder and borderline disorder. There are great variations between the different versions of DSM (DSM-III - DSM-III-R) leading in the case of schizotypal disorder to a reduction of prevalence from 11 per cent to 1 per cent.[69]

On the other hand, normal the population also presents maladaptative variants of traits such as: neuroticism, irritability, vulnerability, anxiety, depression, impulsivity, low consciousness, rush, negligence, hedonism, immorality, unreliable, irresponsibility, high antagonism, manipulation, disappointment, exploitation, aggressiveness, cruelty, heartless.

In everyday practice there is a need for categorical approach, essential to define the frontiers of clinical activity, for research, for management and for forensic questions and also of a dimensional approach, to identify functions and its alterations, relevant for interventions with patients. This could be done investigating in new personality models, adopting the multi-axial version of ICD-10[70] (Table 4.12.1.1.), implementing diagnostic system that include a nosographic approach and idiographic perspective[71] or adopting

Table 4.12.1.1 Multiaxial formulation of ICD-10

Axis I	Clinical disorders (DSM-IV axis I, II and III): personality disorders understood as morbid states of the personality
Axis II	Disablement
Axis III	Relevant non morbid circumstances, including variants of the personality (i.e., Type A behaviour) Normal personality traits (i.e.: five factors model)

Table 4.12.1.2 Proposed five axis system (Charney *et al.* 2002)

Axis I: Genotype	Identification of disease-/ symptom-related genes, of resiliency/protective genes and genes related to the therapeutic response and side effects of psychotropic drugs
Axis II: Neurobiological phenotype	Identification of intermediate phenotypes (neuroimaging, cognitive function, emotional regulation) related to genotype and to targeted pharmacotherapy
Axis III: Behavioural phenotype	Range and frequency expressed behaviours associated with genotype, neurobiological phenotype and environment, related to targeted therapies
Axis IV: Environmental modifiers or precipitators	Environmental factors that alter the behavioural and neurobiological phenotype
Axis V: Therapy	Therapeutic options base don the data of axes I to IV

a new multiaxial system[72] (Table 4.12.1.2). Several authors have claimed for a functional psychopathology. Van Praag[73] has recommended a two-tier diagnosis: 1) nosological diagnosis and 2) psychological dysfunctions correlated with biological variables.

The cosmetic of the personality?

Cosmetic or palliative pharmacotherapy is to use a psychotropic agent to make feel better to a person who is not ill. It usually intents to mitigate unwanted or unaccepted personality traits in order to attain a higher order of social normality and acceptability. Kamer[74] has described how a selective serotonin uptake inhibitor used for the treatment of depression and for other psychiatric alterations can remove personality traits in some people. He has considered traits previously considered as an expression of human misery or, in some cases, as the consequence of negative childhood experiences. Kamer even questions whether there could be a 'pandemic' of cosmetic psychopharmacologies which would lead to the disappearance of phenomena such as anguish which are essential for personal realisation in the arts, religion, and creativity. For example, fluoxetine can make non depressed people feel more vital, mentally more alert and become more popular, leading to an increased feeling of wellness. Shyness became a treatable illness when paroxetine was found to improve the symptoms of social phobia and atomoxetine can change the life conditions of adults with a history of ADHD. Most of these cases are probable subthreshold clinical conditions, something which may lead to a change in diagnostic habits, lowering thresholds or modifying criteria. For instance the DSM criteria of suffering or disablement are value loaded, and new values, such as wellbeing may be introduced.

Conclusions

During the last few decades there has been an impressive growth in research and knowledge on personality disorders. There is a strong growing evidence that they are "real" disorders that can be managed in ways similar to the rest of psychiatric disorders. There is also a growing consensus on the need for a new classification able to capture the nuances beyond the rigidity of present nosological systems. There is also a need for a clearer delimitation from normal personality variants something that will have important impacts, for instance in forensic settings or for reimbursement purposes.

Further information

International Society for the Study of Personality Disorders (ISSPD)
isspd@isspd.com info@isspd.com http://www.isspd.com
International Journal of the ISSPD, Guilford Publications, New York
info@guilford.com

References

1. Thomas, A. and Chess, S.1. (1984) Genesis and Evolution of Behavioral Disorders: From Infancy to Early Adult Life. *Am J Psychiatry*, **141**, 1–9.
2. Kagan, J., Snidman, N., Kahn, V., *et al.* (2007) *The preservation of two infant temperaments into adolescence*. Monographs of the Society for Research in Child Development, Serial No. 287, vol. 72, no. 2. Blackwell, Boston.
3. Foucault, M. (1961) *Folie et déraison. Histoire de la folie à l'âge classique*. Plon, Paris.
4. Pinel, P.H. (1809) *Traité médico-philosophique sur l'alienation mentale*. Brosson, Paris.
5. Cleckley, H. (1941) *The mask of sanity*. Henry Kimpton, London.
6. Prichard, J.C. (1835) *A treatise on insanity and other disorders affecting the mind*. Sherwood, Gilbert and Piper, London.
7. Morel, B. (1859) *Traité des dégénéréscences physiques, intellectuelles et morales de l'espèce humaine*. Paris.
8. Tyrer, P., Casey, P., and Ferguson, B. (1991) Personality disorder in perspective. *British Journal of Psychiatry,* **159**, 463–71.
9. Cusack, J.R. and Malaney, K.R. (1992). Patients with antisocial personality disorder. Are they bad or mad? *Postgraduate Medicine*, **91**, 341–4, 349–52, 355.
10. Maden,T. and Tyrer, P. (2003) Dangerous and severe personality disorders: a new personality concept from the United Kingdom. *Journal of Personality Disorders*, **17**, 489–96.
11. Tyrer, P. (2007) An agitation of contrary opinions. *British Journal of Psychiatry*, **190**, (suppl.49), 1–2.
12. Griesinger, W. (1872) *Gesammelte Abhandlungen. Vol. I, Psychiatrische und nervenpathologische Abhandlungen.* Reprinted by Bonset, Amsterdam, 1968.
13. Schneider, K. (1971) *Klinische Psychopathologie* (9th edn). Thieme, Stuttgart.
14. Schneider, K. (1950) *Die psychopathischen Persönlichkeiten* (9th revised edn). Deuticke, Vienna.
15. López Ibor, J.J. (1966) *Las neurosis como enfermedades del ánimo. Gredos, Madrid.*
16. American Psychiatric Association (1994*) Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Association, Washington, DC.
17. World Health Organization (1992*) The ICD-10 classification of mental and behavioural disorders—clinical descriptions and diagnostic guidelines*. World Health Organization, Geneva.
18. Schwartz, M.A. and Wiggins, O.P. (1987) Diagnosis and ideal types: a contribution to psychiatric classification. *Comprehensive Psychiatry*, **28**, 277-91.
19. Personal communication
20. Pervin, L. (Ed.) (1990) *Handbook of personality: Theory and research*. The Guilford Press. New York.
21. Freud, S., Freud, A., Strachey, J. (1953–1974) *The standard edition of the complete psychological works of Sigmund Freud*. Hogarth Press Richmond UK.
22. Jung, C.J. (1966) *Two essays on analytical psychology*. Princeton University Press, Princeton, NJ.

23. Eysenck, H.J. (1970) *The structure of human personality*. Methuen, London.
24. Eysenck, H.J. and Eysenck, M.W. (1985*) Personality and individual differences: A natural science approach*. Plenum Press, New York.
25. Zuckermann, M. (1979) *Sensation seeking: Beyond the optimal level of arousal*. Erlbaum, Hillsdale, NJ.
26. Oreland, L., Wiberg, A., Åsberg, M., *et al.* (1981) Platelet MAO activity general clinical description of personality disorders and monoamine metabolism in CSF in depressed and suicidal patients and in healthy controls. *Psychiatric Research*, **4**, 21–9.
27. Siever, L.J. and Davis, K.L. (1991) A psychobiological perspective on the personality disorders. *American Journal of Psychiatry*, **148**, 1647–58.
28. Cloninger, C.R. (1987) A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry*, **44**, 573–88.
29. Cloninger, C.R. (1996) A psychobiological model of temperament and character. Fundamental findings for use in clinical practice. In *New research in psychiatry* (ed. H. Häfner and E.M. Wolpert), pp. 95–112. *Hogrefe and Huber, Göttingen.*
30. McCrae, R.R. and John, O.P. (1992) An introduction to the five-factor model and its applications. *Journal of Personality*, **60**, 175–213.
31. Loehlin, J.C. (1992) *Genes and environment in personality development*. Sage, Newbury Park, CA.
32. Digman, J.M. (1997) Higher-order factors of the Big Five. *Journal Personality and Social Psychology*, **73**, 1246–56.
33. Ashton, M.C. and Lee, K. (2005) Honesty-humility, the big five, and the five-factor model. *Journal of Personality*, **73**, 1321–53.
34. Church, A.T., Reyes, J.A., Katigbak, M.S., *et al.* (1997) Filipino personality structure and the big five model: a lexical approach. *Journal of Personality*, **65**, 477–528.
35. Haas, H.A. (2002) Extending the search for folk personality constructs: the dimensionality of the personality-relevant proverb domain. *Journal of Personality and Social Psychology*, **82**, 594–609.
36. Luminet, O., Bagby, R.M., Wagner, H., *et al.* (1999) Relation between alexithymia and the five-factor model of personality: a facet-level analysis. *Journal of Personality Assessment*, **73**, 345–58.
37. Jaspers, K. (1909) *Heimweh und Verbrechen*. F.C.M. Vogel, Leipzig.
38. Durkheim, E. (1951) *Suicide*. Free Press, Glencoe, IL.
39. Kraus, A. (1977) *Soziales Verhalten und Psychosen manisch-depressiver*. Enke, Stuttgart.
40. López-Ibor, J.J. (1991) Obsessive-compulsive disorder and other disorders. *European Neuropsychopharmacology*, **1**, 275–9.
41. Lacey, J.H. and Evans, C.D.H. (1986). The impulsivist: a multi-impulsive personality disorder. *British Journal of Addiction*, **81**, 641–9.
42. Janet, P. (1903) *Les obsessions et la psychasthenie*. Alcan, Paris.
43. Widiger, TA. (2007) Dimensional models of personality disorder. *World Psychiatry,* **6**, 15–9.
44. Deary, I.J., Peter, A., Austin, E., *et al.* (1998) Personality traits and personality disorders. *British Journal of Psychology*, **89**, 647–61.
45. Livesley, W.J., Jang, K.L., Vernon, P.A. (1998) Phenotypic and genetic structure of traits delineating personality disorder. *Archives of General Psychiatry*, **55**, 941–8.
46. Livesley, W.J. (1998) Suggestions for a framework for an empirically based classification of personality disorder. *Canadian Journal of Psychiatry*, **43**, 137–147.
47. Markon, K.E., Krueger, R.F., Watson, D. (2005) Delineating the structure of normal and abnormal personality: an integrative hierarchical approach. *Journal of Personality and Social Psychology*, **88**, 139–57.
48. O'Connor, B.P. (2002) The search for dimensional structure differences between normality and abnormality.. *Journal of Personality and Social Psychology*, **83**, 962–82.
49. Trull, T.J. (1992) DSM-III-R personality disorders and the five-factor model of personality: An empirical comparison. *Journal of Abnormal Psychology*, **101**, 553–60.
50. Blais, M.A. (1997) Clinician ratings of the five-factor model of personality and the DSM-IV personality disorders. *Journal of Nervous and Mental Diseases*, **185**, 388–93.
51. Parker, G., Hadzi-Pavloic, D., Wilhelm, K. (2000) Modeling and measuring the personality disorders. *Journal of Personality Disorders*, **14**, 189–98.
52. Mulder, R.T., Joyce, P.R. (1997) Temperament and the structure of personality disorder symptoms. *Psychologial Medicine*, **27**, 99–106.
53. Lynam, D.R. (1996) Early identification of chronic offenders: who is the fledgling psychopath? *Psychological Bulletin*, **120**, 209–34.
54. Seagrave, D., Grisso, T. (2002) Adolescent development and the measurement of juvenile psychopathy. *Law and human behavior*, **26**, 219–39.
55. Hart, S.D., Watt, K.A., Vincent, G.M. (2002) Commentary on Seagrave and Grisso: impressions of the state of the art. *Law and human behavior*, **26**, 241–5.
56. Widiger, T.A., Frances, A. (1985) Axis II personality disorders: diagnostic and treatment issues. *Hospital and Community Psychiatry*, **36**, 619–27.
57. Akiskal, H.S., Akiskal, K.K., Perugi, G., *et al.* (2006) Bipolar II and anxious reactive 'comorbidity': Toward better phenotypic characterization suitable for genotyping. *Journal of Affective Disorders*, **96**, 239–47.
58. Clark, L.A. (2005) Temperament as a unifying basis for personality and psychopathology. *Journal of Abnormal Psychology*, **114**, 505–21.
59. Ricciardi, J.N., Baer, L., Jenike, M.A., *et al.* (1992) Changes in DSM-III-R axis II diagnoses following treatment of obsessive-compulsive disorder. *American Journal of Psychiatry*, **149**, 829–31.
60. Rosenthal, D. (1970) *Genetic theory and abnormal behavior*. New York: McGraw Hill.
61. Tyrer, P. (2007) Personality diatheses: a superior explanation than disorder. *Psychological Medicine*, **12**, 1–5.
62. López-Ibor, J.J., Jr, Lana, F., and Sáiz, J. (1991) *Serotonin, impulsiveness and aggression in humans*. In Serotonin related psychiatric syndromes: clinical and therapeutic links (ed. G.B. Cassano and H.S. Akiskal), pp. 35–9. Royal Society of Medicine, London.
63. Sapolsky, R.M. (1990) Adrenocortical function, social rank, and personality among wild baboons. *Biological Psychiatry*, **15**, 862–78.
64. Manzano, J., Zabala, I., Borella, E., *et al.* (1992) Continuity and discontinuity of psychopathology: a study of patients examined as children and as adults. I. Antecedents of adult schizophrenic disorders. *Schweierz Archiv für Neurologie und Psychiatrie,* **143**, 5–25.
65. Lebow, J. and Gordon, K.C. (2006) You cannot choose what is not on the menu—obstacles to and reasons for the inclusion of relational processes in the DSM-V: comment on the special section. *Journal of Family Psychology*, **20**, 432–7.
66. Wakefield, J.C. (2006) Are there relational disorders? A harmful dysfunction perspective: comment on the special section. *Journal of Family Psychology*, **20**, 423–7.
67. Erickson, T.M. and Pincus, A.L. (2005) Using Structural Analysis of Social Behavior (SASB) measures of self- and social perception to give interpersonal meaning to symptoms: anxiety as an exemplar. *Assessment*, **12**, 243–54.
68. Beach, S.R., Wamboldt, M.Z., Kaslow, N.J., *et al.* (Eds.) (2006) *Relational Processes and DSM-V. Neuroscience, Assessment, Prevention, and Treatment*. American Psychiatric Publishing Inc., Washington.
69. World Health Organization (1992) *The ICD-10 International Classification of Mental and Behavioral Disorders. Clinical descriptions and diagnostic guidelines*. Geneva, WHO.
70. Widiger, T.A., Costa, P.T. Jr. (1994) Personality and personality disorders. *Journal of Abnormal Psychology*, **103**, 78–91.

71. Blashfield, R., Blum, N., Pfohl, B. (1992) The effects of changing axis II diagnostic criteria. *Comprehensive Psychiatry*, **33**, 245–5.
72. Janca, A., Kastrup, M., Katschnig, H., *et al.* (1996) The ICD-10 Multiaxial System for Use in Adult Psychiatry: Structure and Applications. *The Journal of Nervous and Mental Diseases*, **184**, 191–192.
73. Mezzich, J.E. (2007) Psychiatry for the Person: articulating medicine's science and humanism. *World Psychiatry*, **6**, 1–3.
74. Charney, D.S., Barlow, D.H., Botteron, K., *et al.* (2002) *Neuroscience research agenda to guide development of a pathophysiologically based classification system. In: A Research Agenda for DSM-V. Kupfer, D., First, M., Regier, D.A. (Eds.), pp:31–83*, American Psychiatric Association, Washington.
75. Van Praag, H.M. (1990). Two-tier diagnosing in psychiatry. *Psychiatry Research*, **34**, 1–11.
76. Kamer, P.D. (1993) *Listening to Prozac*. Viking, New York.

4.12.2 Diagnosis and classification of personality disorders

James Reich and Giovanni de Girolamo

Definitions of personality disorders

There has been considerable interest in the study of personality and personality disorder (PD) since early times and in many different cultures. However, as noted by Tyrer *et al.*[1] 'The categorization of personality disorder did not receive any firm support until the time of Schneider'. Schneider[2] regarded abnormal personalities as 'constitutional variants that are highly influenced by personal experiences' and identified 10 specific types or classes of 'psychopathic personality'. The classification system proposed by Schneider has deeply influenced subsequent classification systems[1]: of the 10 types of PD identified by Schneider, eight are closely related to similar types of PD as classified in DSM-III.[3] Many of these categories are also represented in DSM-IV[4] and ICD-10.[5]

Personality is defined in the second edition of the WHO *Lexicon of Psychiatric and Mental Health Terms*[6] as 'The ingrained patterns of thought, feeling, and behaviour characterising an individual's unique lifestyle and mode of adaptation, and resulting from constitutional factors, development, and social experience'. Personality disorders, according to the ICD-10 diagnostic guidelines[5]:

> . . . comprise deeply ingrained and enduring behaviour patterns, manifesting themselves as inflexible responses to a broad range of personal and social situations. They represent either extreme or significant deviations from the way the average individual in a given culture perceives, thinks, feels, and particularly, relates to others. Such behaviour patterns tend to be stable and to encompass multiple domains of behaviour and psychological functioning. They are frequently, but not always, associated with various degrees of subjective distress and problems in social functioning and performance.

For example, a dependent PD in a favourable environment might not cause dysfunction, but nevertheless might be considered a disorder since it is clinically identical to the same disorder that usually causes dysfunction.

DSM-IV[4] defines a PD as 'an enduring pattern of inner experience and behaviour that deviates markedly from the expectations of the individual's culture'. The pattern is manifested in two or more of the following areas: cognition, affectivity, interpersonal functioning, and impulse control. The pattern is inflexible and pervasive across a broad range of situations, has an early onset, is stable and leads to significant distress or impairment.

Personality traits, according to DSM-IV,[4] 'are enduring patterns of perceiving, relating to and thinking about the environment and oneself that are exhibited in a wide range of social and personal contexts. Only when personality traits are inflexible and maladaptive and cause significant functional impairment or subjective distress do they constitute PDs.'

ICD and DSM classifications of personality disorders

Table 4.12.2.1 lists the specific PDs as classified in ICD-9,[7] ICD-10, DSM-IIIR,[8] and DSM-IV.

In the ICD-10 classification, which does not have a multiaxial system for the separate recording of the personality status, PD can be diagnosed together with any other mental disorder, if present. Although a multiaxial system for ICD-10 is being developed, this will not include a separate axis for PDs, as in DSM-IV.

Despite the importance given to behavioural manifestations for the classification and assessment of PDs, personality traits and attitudes are also considered when a diagnosis is made. The ICD-10 diagnostic guidelines subdivide PDs 'according to clusters of traits that correspond to the most frequent or conspicuous behavioural manifestations'. As stressed by Widiger and Frances,[9] the reliance on behavioural indicators can improve inter-rater reliability, which reduces the amount of inferential judgement required for the diagnosis, but it does not ensure that the same diagnosis will be made at different times. Moreover, the diagnosis of a PD cannot be based on a single behaviour, as any given behaviour may have multiple causes (e.g. situational and role factors).

There have been four studies that have explored the diagnostic categories for PDs contained in ICD-10 and compared them with the DSM classification. The first[10] was carried out among 177 American clinicians who found some degree of overlap between the different categories. When the authors compared the diagnostic categories in ICD-10 with those in DSM-IIIR, they found that only anankastic (ICD) and obsessive–compulsive (DSM) PDs showed a high level of correspondence. The second study[11] looked at 52 outpatients and compared DSM-IIIR to ICD-10. It found fair concordance for the diagnosis of 'any PD', but poor agreement for individual PDs; the ICD-10 tended to overdiagnose PDs relative to DSM-IIIR. The third report[12] compared ICD-10 and DSM-IV in 58 patients with panic disorder. There was good agreement for the presence of 'any PD', and a reasonable agreement between individual diagnoses (κ ranged from 0.51 to 0.83.), with a tendency for ICD-10 to overdiagnose PDs relative to DSM-IV. In the fourth study,[13] ICD-10 criteria were found to have satisfactory inter-rater reliability in a sample of homeless adults.

In the American taxonomic system, a multiaxial classification was first introduced in DSM-III. With the development of DSM-IIIR, more than 100 changes in the classification of PDs were introduced compared with DSM-III.[14,15] While the multiaxial and categorical style of classification was maintained, the diagnostic criteria were revised to form a list of symptoms for each PD, of

Table 4.12.2.1 Comparison of different classification systems of personality disorders: ICD-9, ICD-10, DSM-IIIR, and DSM-IV

ICD-9	ICD-10	DSM-III-R	DSM-IV
Paranoid personality disorder	Paranoid personality disorder	Paranoid personality disorder	Paranoid personality disorder
Schizoid personality disorder	Schizoid personality disorder	Schizoid personality disorder	Schizoid personality disorder
Personality disorder with predominantly sociopathic or asocial manifestations	Dissocial personality disorder	Antisocial personality disorder	Antisocial personality disorder
	Emotionally unstable personality disorder:		
Explosive personality disorder	Impulsive type	NA	NA
NA	Borderline type	Borderline personality disorder	Borderline personality disorder
Histrionic personality disorder	Histrionic personality disorder	Histrionic personality disorder	Histrionic personality disorder
Anankastic personality disorder	Anankastic personality disorder	Obsessive-compulsive personality disorder	Obsessive-compulsive personality disorder
NA	Anxious [avoidant] personality disorder	Avoidant personality disorder	Avoidant personality disorder
NA	Dependent personality disorder	Dependent personality disorder	Dependent personality disorder
Affective personality disorder	Other specific personality disorders	Passive-aggressive personality disorder	NA
Asthenic personality disorder		Schizotypal personality disorder	Schizotypal personality disorder
		Narcissistic personality disorder	Narcissistic personality disorder
		Self-defeating personality disorder	NA
		Sadistic personality disorder	NA
			Personality disorder not otherwise specified

which only a certain number were required for a diagnosis to be reached. In DSM-IIIR, each category of PD comprised 7 to 10 criteria, with the presence of four to six criteria required for diagnosis. DSM-IIIR contained 11 PDs (see Table 4.12.2.1), plus two new disorders (self-defeating PD and sadistic PD) that were not included in DSM-III but were considered as diagnostic categories needing further study. As in DSM-III, the 11 PDs were divided into three clusters:

- cluster A (the 'odd' or 'eccentric' cluster), which included paranoid, schizoid, and schizotypal PD;
- cluster B (the 'dramatic' or 'erratic' cluster), which included histrionic, narcissistic, antisocial, and borderline PDs; and
- cluster C (the 'anxious' cluster), which included avoidant, dependent, obsessive–compulsive, and passive–aggressive PDs.

One study in the United States examined changes in personality diagnoses using DSM-III versus DSM-IIIR.[16] For some categories there was a considerable difference in the frequency of diagnosis: for example, there was an 800 per cent increase in the rate of schizoid PD and a 350 per cent increase in the rate of narcissistic PD diagnosed by the clinicians when DSM-IIIR criteria were applied.

DSM-IV was designed to be a conservative evolution from DSM-IIIR; however, some differences in diagnoses between DSM-IIIR and DSM-IV can be expected.[17] In general, the different DSMs should not be considered interchangeable unless there is specific data supporting agreement of a diagnosis across systems. As shown in Table 4.12.2.1, DSM-IV includes 11 PDs as in the DSM-IIIR classification; slight changes were introduced in the diagnostic criteria, and a new category 'PD not otherwise specified' added. Passive–aggressive, self-defeating, and sadistic PDs (provisionally included in DSM-IIIR) were dropped. The overall effect of these changes will be to increase the concordance between the DSM-IV and the ICD-10 classification systems compared with that between DSM-IIIR and ICD-10. DSM-IV also includes the three clusters present in DSM-IIIR.

Similarities differences between ICD-10 and DSM-IV

Table 4.12.2.1 shows that for seven categories of PD (paranoid, schizoid, dissocial/antisocial, histrionic, anankastic/obsessive–compulsive, anxious/avoidant, and dependent), there is a specific correspondence between ICD-10 and DSM-IV. For three categories, there are differences in nomenclature between the two systems; in particular ICD-10 uses the term 'anankastic' instead of 'obsessive–compulsive', to avoid the erroneous implication of an inevitable link between this type of personality and obsessive–compulsive disorder. ICD-10 also uses the term 'dissocial' instead of 'antisocial', to prevent any possible connotation of stigmatization, and the term 'anxious' instead of 'avoidant'. Moreover, while DSM-IV classifies borderline PD as a specific category, ICD-10 includes it as a subcategory of emotionally unstable PD. Narcissistic and passive–aggressive PDs (present in DSM-IV) are included in ICD-10 under the category of 'other specific PDs'. Finally, while DSM-IV includes schizotypal PD as a PD, ICD-10 classifies it in the overall group of 'Schizophrenia, schizotypal and delusional disorders', to highlight the contiguity between this disorder and the schizophrenia group disorders, as shown by genetic and clinical studies. DSM-IV has the category 'Personality disorders not otherwise

specified', while ICD-10 has the category 'Other specific personality disorders'.

Changes in the conceptualization of DSM personality disorders since the last edition of this chapter

Empirical research has advanced in the years following the original chapter in an earlier volume. These changes have impacted our understanding and use of DSM measurement instruments. These changes are that the personality disorders as described by DSM are not as enduring as we once thought. The instruments to measure the DSM PDs have modest agreement at best on the categorical level. Finally these instruments do not seem to adequately fit most of the disorders diagnosed which are diagnosed in the remainder category, 'Personality Disorder NOS'.

1. Research indicating lack of enduring quality of personality disorders.

There has now been considerable research indicating that some aspects of personality are state like. This was a line of research pursued by Reich[18,19] and later confirmed by others.[20] This means that some personality traits may disappear relatively rapidly—the state component. Experienced researchers have also found that even when personality disorders are selected for long-term study by careful methods, significant percentages of these will not be found on retest within a 6 month to several year periods.[21,22]

2. Modest agreement of personality DSM personality disorder instruments.

Research has been fairly consistent from the beginning of DSM instruments that while they may measure aspects of clinical relevance even well-designed instruments of the same design (self-report or semi-structured interview) are compared on categorical diagnoses, the agreements are usually modest at best.[23] (The results improve somewhat when using dimensional measures.) In addition when personality disorders are measured in a clinical population very few fit into the established categories and most wind up in the 'remainder bin' of personality disorder NOS.[24]

3. The problem of chasing changing criteria.

Each time the wording of a questionnaire is changed it can make significant difference in the outcome of the questionnaire or survey even though the changes may seem minor. This means that for every change in a version of the DSM, the DSM personality test developers would theoretically have to do extensive reliability and validity testing all over again. Unfortunately this is a time-consuming and expensive process which is beyond the resources of most DSM test developers. What happens instead is that either they do not update their instrument or update only the questions to conform to the new DSM criteria without doing new validity testing. Although these updated tests may have 'face validity' they (understandably) do not have extensive reliability and validity testing.

4. Conclusions from intervening research on DSM personality instruments.

At the end of the day we are left with the DSM personality measurement instruments being limited by the conceptualizations behind them which may be incorrect. These errors may either by in the enduring nature of the personality disorders and the nature of the proper categories or the nature of the specific criteria for a category. These measures then become good dimensional measures of various aspects of personality pathology and a rough clinical guide for 'disorder'. However, no one should expect that what one of these instruments is measuring is really the same as another. They do, however, conform to the DSM nomenclature.

Categorical versus dimensional styles of classification

In general, researchers involved in the assessment of personality traits tend to use dimensional measures based on normal populations. In contrast, those concerned with personality types and, even more, clinicians concerned with PDs, tend to employ categorical concepts and assessment measures based on these concepts.[15]

Each of the two approaches has specific advantages and disadvantages. The drawbacks of the categorical approach are represented by the difficulty of classifying patients who are at the boundary of different categories or who do not meet the diagnostic criteria for any specific PD, but who still have significant pathology. Other points that should be addressed include the wide variation of symptomatology found within each given category, the need for heterogeneous categories, such as 'mixed' and 'atypical', the need to simplify necessarily complex conditions, the need to define valid cut-off points, and the use of a nominal rather than an ordinal scale.

Those in favour of a dimensional approach argue that PDs differ from normal variations in personality only in terms of degree, and to some extent this is supported by empirical data.[25,26] However, there is evidence that some dimensional models do not account for all of the abnormal personality traits (see discussion on the 'Big five' models below.)

Moreover, it is still unclear whether normal and abnormal personality traits are the same or whether they are qualitatively different. Two main findings seem to support the latter hypothesis: first, normal personality traits are at least moderately heritable, while abnormal personality traits appear less heritable; second, the prevalence rates of PDs found in surveys are much greater than would be expected from the prevalence rates of normal personality traits in the general population. Some researchers[27] have even suggested that an extreme form of a normal personality trait is not necessarily pathological. Although some authors argue that there is no difference between normal and abnormal personality traits,[25] and this is still an area of open inquiry, it does appear more and more likely that there are some abnormal personality traits that differ from normal ones. It is possible that different models may be appropriate in different situations.

1. 'Big Five' personality measures and DSM-IV personality disorders

Since the first edition of this chapter there has been a fair amount of research using the 'big five' personality factors. There is evidence that the big five factors can account for some, but not all of the pathological personality pathology described by the DSM system.[28] However the amount of information also depends on the specific instrument with some with more facets or detail being more able account for aspects of pathological personality. These would be instruments such as the NEO-PI[29] and the five factor model of Cloninger.[30] Also of possible greater utility are dimensional instruments which were designed from the start to measure

abnormal as well as normal personality. These include the Schedule for Non-adaptive and Adaptive Personality (SNAP)[31] and Dimensional Assessment of Personality Pathology (DAPP).[32] Also worth noting is a method by Widiger which combines a five factor self-report with a five factor interview in order to obtain more comprehensive personality information.[33]

2. Other non-DSM personality measures from which DSM PD diagnoses can be derived.

There is one measure of DSM-IV personality which is derived from the interpersonal circumplex model. This is the Wisconsin Personality Disorders Inventory (currently in version IV).[34] Although from a fairly different theoretical perspective from the DSM-IV it does have translations to make some DSM-IV diagnoses.[35]

Another unique approach is the Shedler–Westen Assessment Procedure which does not have a patient interview but rather has a clinician who knows the patient do a Q sort procedure. The process is based on distinguishing personality disorders based on prototypes. This procedure can be translated into DSM personality diagnoses. It tends to create fewer diagnoses because the number of descriptors that can be used is limited by the Q sort procedure.[36]

The Millon Clinical MultiAxial Inventory (now in version three) is a very commonly used personality measure. It is now based on evolutionary theory according to its manual. One of the best validated instruments in data-based terms. Unfortunately this data is mostly published in its own manual and not the journal literature. As the population that the instrument was validated on included some fairly severe disorders it is not clear how well the findings would translate into a less severely ill population.[37]

Assessment methods for personality disorders

Table 4.12.2.2 shows the main methods currently available for assessing all PDs. Additional instruments for assessing specified PDs have also been developed—some of the methods listed are

Table 4.12.2.2 Some commonly used assessment methods for all DSM personality disorders

Name of the instrument	Author(s) [a]	Method of assessment	Number of Questions	Time required (minutes)
Diagnostic Interview for Personality Disorders (DIPD)	Zanarini	Semistructured interview with patient using DSM-IV criteria	398	60–120
Dimensional Assessment of Personality Disorders (DAPP-BQ)	Livesley & Jackson	Dimensional based of normal and abnormal personality. Has components of DSM-IV PDs	290	45
International Personality Disorder Examination (IPDE)	Loranger *et al.*	Semistructured interview with patient using ICD-10 and DSM-IV criteria	537	150
Millon clinical Multiaxial Inventory (MCMI)	Millon	Self-report by patient using DSM-IV criteria	175	20–30
NEO Personality Inventory-Revised (NEO PI-R)	Costa & McRae	Comprehensive measurement of normal and abnormal personality traits.	240	30–40
Personality Assessment Schedule (PAS)	Tyrer *et al.*	Semistructured interview with informant (s) can derive ICD-10 and DSM-IV diagnoses.	24	60
Personality Diagnostic Questionnaire –Revised (PDQ-4)	Hyler *et al.*	Self-report by patient or informant(s) using DSM-IV criteria	99	30
Personality Disorders Interview-IV	Widiger *et al.*	Semistructured interview with patient using DSM-IV criteria	325	60–120
Schedule for Nonadaptive and Adaptive Personality (SNAP)	Clark	Self report by patient using DSM-IV and dimensional criteria	375	30–60
Shedler-Westen Assessment Procedure	Westen & Shedler	Clinician rated Q sort procedure. Gives abnormal traits, symptoms and defences and can generate DSM-IV diagnoses	200	No interview, based of clinician knowledge of client.
Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II)	First & Gibbon	Semistructured interview with patient using DSM-IV criteria	303	60–90
Structured Interview for DSM Personality Disorders-IV (SIPD)	Pfohl *et al.*	Semistructured interview with patient or informant(s) using DSM-III criteria	337	90
Temperament and Character Inventory (TPQ)	Cloninger *et al.*	Self-report by patient. Temperament and Character dimensions.	240	30–40
Wisconsin Personality Disorders Inventory (WISPI)	Klein *et al.*	Self-report by patient using DSM-III-R criteria. Correlated with interpersonal object relation theory.	214	30

[a] For specific references to each instrument please see references.[30, 31, and 33]

new, while others have been revised two or three times.[38,39] The following points related to these methods need to be mentioned.

1 The interview measures have generally shown a satisfactory inter-rater reliability, while test–retest reliability has not been well established. However, three methods do show some evidence of good test–retest reliability—the Personality Assessment Schedule (**PAS**),[40] the International Personality Disorder Examination (**IPDE**),[41] and the Structured Clinical Interview for Personality Disorders (**SCID-II**).[42] Many of the methods have been standardized on psychiatric inpatient or outpatient populations; their applicability in epidemiological community studies is largely unknown.[43,44]

2 The various measures tend not to agree with each other on specific diagnoses.[38,39] A measurement on one standardized instrument is not necessarily the same as the measurement on another.

3 Some authors, mostly developers of interview instruments, have in the past stated that self-report measures are not as valid for the measurement and study of personality as interview measures. These arguments tend to be based on the finding that self-report instruments do not agree well with interview instruments, and that a PD diagnosis cannot be made without a clinical interview. The first argument does not hold water, since none of the interview instruments agree well with each other either. Whether self-report instruments can reliably diagnose a PD (as opposed to personality traits) is an open question. However, dimensional interview instruments have high test–retest reliability (some for as long as 30 years) and have been, and will continue to be, a valuable component of personality research, especially where the focus is on dimensions.[45] Self-report instruments measuring DSM disorders tend to disagree with each other in a similar way to the interview instruments. In conclusion, the instrument chosen for any given clinical or research endeavour should reflect the ability of the individual instrument to meet the specific needs of the project. Most researchers now believe that using a self-report, which may be more sensitive but less specific is a good prelude to a semi-structured interview measure.

4 Standardized testing of personality and clinician impression do not tend be in good agreement. This is due both to the tendency of instruments to report more disorders and for clinicians to use idiosyncratic criteria in their diagnoses.[46,47] Standardized measures are a must, however, if the data is to be used for research or public policy purposes.

5 Most personality measurement instruments are affected by the comorbid presence of a non-personality emotional disorder (Axis I disorder in the DSM system). Some structured interviews try to correct for this by asking patients about times when they were not suffering from an Axis I disorder; however, when the Axis I disorder is chronic, this may be difficult to achieve.[41,42] This problem also affects the ability of questionnaires to differentiate between current Axis I disorders and PDs, as this self-judgement of patients who are suffering from a psychiatric condition is frequently impaired.[41] The PDE, PAS, and SCID-II may be less affected by this problem.

6 There is disagreement among experts about the use of informants. Many authors have argued that, besides the patient, a key informant should also be interviewed, given the likelihood that many patients will not reply reliably to questions about their personality and the possibility that informant ratings will differ substantially from patient ones.[41,48] However, even if an informant is interviewed, it is often unclear which source to use to score the test if the interviewee and informant disagree. This tends to reduce their value. More recent research indicates that informants may be of value if their information is in certain areas that are easily observable to then and then are incorporated into the evaluation process in a standardized way.[49]

7 Discriminant validity refers to the ability of a diagnostic system and measurement system to diagnose non-overlapping disorders. Discriminant validity is not high with the ICD-10 and DSM-IV PDs.[50] This means that it is the rule, rather the exception, that multiple personality diagnoses will be made: some studies have provided evidence of this assumption. For instance, in four studies that examined personality comorbidity in a total of 568 patients, the average percentage of multiple diagnoses was 85 per cent.[51] To what extent this reflects the real coexistence of different disorders—with distinct patterns of symptoms, correlates, and course—or if it is simply the effect of an insufficient discriminant validity of current diagnostic systems has still to be clarified.

Further information

Reich, J. (2005). State and trait in personality disorders. In *Personality disorders: current research and treatments* (ed. J. Reich), pp. 3–20. Taylor & Francis, New York.

Reich, J. (2007). State and trait in personality disorders. *Annals of Clinical Psychiatry*, **19**, 37–44.

Widiger, T.A., Costa, P.T., and Samuel, D.B. (2006). Assessment of maladaptive personality traits. In *Differentiating normal and abnormal personality* (2nd edn) (eds. S. Strack and M. Lorr), pp. 311–55. Springer, New York.

References

1. Tyrer, P., Casey, P., and Ferguson, B. (1991). Personality disorder in perspective. *British Journal of Psychiatry*, **159**, 463–71.
2. Schneider, K. (1923). *Die psychopathischen personlichkeiten*. Springer, Berlin.
3. American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Press, Washington, DC.
4. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
5. World Health Organization. (1992). *International statistical classification of diseases and related health problems, 10th revision*. WHO, Geneva.
6. World Health Organization. (1994). *Lexicon of psychiatric and mental health terms* (2nd edn). WHO, Geneva.
7. World Health Organization. (1977). *International classification of diseases, injuries and causes of death, ninth revision*. WHO, Geneva.
8. American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd edn, revised). American Psychiatric Press, Washington, DC.
9. Widiger, T.A. and Frances, A. (1985). Axis II personality disorders: diagnostic and treatment issues. *Hospital and Community Psychiatry*, **36**, 619–27.

10. Blashfield, R.K. (1991). An American view of the ICD-10 personality disorders. *Acta Psychiatrica Scandinavica*, **82**, 250–6.
11. Sara, G., Raven, P., and Mann, A. (1996). A comparison of DSM-III-R and ICD-10 personality disorder criteria in an outpatient population. *Psychological Medicine*, **26**, 151–60.
12. Starcevic, V., Bogojevic, G., and Kelin, K. (1997). Diagnostic agreement between the DSM-IV and ICD-10-DCR personality disorders. *Psychopathology*, **30**, 328–34.
13. Merson, S., Tyrer, P., Duke, P., *et al.* (1994). Interrater reliability of ICD-10 guidelines for the diagnosis of personality disorders. *Journal of Personality Disorders*, **8**, 89–95.
14. Gorton, G. and Akhtar, S. (1990). The literature on personality disorders, 1985–1988: trends, issues and controversies. *Hospital and Community Psychiatry*, **41**, 39–51.
15. Widiger, T.A., Frances, A., Spitzer, R.L., *et al.* (1988). DSM-III-R Personality disorders: an overview. *American Journal of Psychiatry*, **145**, 786–95.
16. Morey, L.C. (1988). Personality disorders in DSM-III and DSM-III-R: convergence, coverage and internal consistency. *American Journal of Psychiatry*, **145**, 573–7.
17. Coolidge, F.L. and Segal, D.L. (1998). Evolution of personality disorder diagnosis in the diagnostic and statistical manual of mental disorders. *Clinical Psychology Review*, **19**, 583–99.
18. Reich, J. (2005). State and trait in personality disorders. In *Personality disorders: current research and treatments* (ed. J. Reich), pp. 3–20. Taylor & Francis, New York.
19. Reich, J. (2007). State and trait in personality disorders. *Annals of Clinical Psychiatry*, **19**, 37–44.
20. Clark, L.A. (2005). Stability and change in personality pathology: revelations of three longitudinal studies. *Journal of Personality Disorders*, **19**, 524–32.
21. Zanarini, M.C., Frankenburg, F.R., Hennen, J., *et al.* (2006). Prediction of the 10-year course of borderline personality disorder. *American Journal of Psychiatry*, **163**, 827–32.
22. Skodol, A.E., Gunderson, J.G., Shea, M.T., *et al.* (2005). The collaborative longitudinal personality disorders study (CLPS). *Journal of Personality Disorders*, **19**, 487–504.
23. Widiger, T.A. and Coker, L.A. (2002). Assessing personality disorders. In *Clinical personality assessment. Practical approaches* (2nd edn) (ed. J.N. Butcher), pp. 407–34. Oxford University Press, New York.
24. Verheuk, R. and Widiger, T.A. (2004). A meta-analysis of the prevalence and usage of the personality disorders not otherwise specified (PDNOS) diagnosis. *Journal of Personality Disorders*, **18**, 309–19.
25. Livesley, W.J., Schroeder, M.L., Jackson, D.N., and Jang, K.L. (1994). Categorical distinctions in the study of personality disorders: implications for classification. *Journal of Abnormal Psychology*, **103**, 6–17.
26. Schroeder, M.L. and Livesley, W.J. (1991). An evaluation of DSM-III-R personality disorders. *Acta Psychiatrica Scandinavica*, **84**, 512–59.
27. Birtchnell, J. (1988). Defining dependence. *British Journal of Medical Psychology*, **61**, 111–23.
28. Clark, L.A. and Harrison, J.A. (2001). Assessment instruments. In *Handbook of personality disorders: theory, research and treatment* (ed. W.J. Livesley), pp. 277–306. Guilford Press, New York.
29. Costa, P.T. and McCrae, R.R. (1992). *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI) professional manual*. Psychological Assessment Resources, Odessa, FL.
30. Cloninger, C.R., Przybeck, T.R., Svrakic, D.M., *et al.* (1994). *The temperament and character inventory (TCI): a guide to its development and use*. Center for Psychobiology of Personality, Washington University, St. Louis, MO.
31. Clark, L.A. (1993). *Schedule for adaptive and nonadaptive personality*. University of Minnesota Press, Minneapolis, MN.
32. Livesley, W.J. and Jackson, D.N. (2006). *Manual for the dimensional assessment of personality pathology*. Sigma, Port Huron, MI.
33. Widiger, T.A., Costa, P.T., and Samuel, D.B. (2006). Assessment of maladaptive personality traits. In *Differentiating normal and abnormal personality* (2nd edn) (eds. S. Strack and M. Lorr), pp. 311–55. Springer, New York.
34. Klein, M.H. and Benjamin, L.S. (1996). *The Wisconsin Personality Inventory-IV*. Wisconsin Psychiatric Institute, Madison, WI.
35. Smith, T.L., Klein, M.H., and Benjamin, L.S. (2003). Validation of the Wisconsin Personality Inventory-IV with the SCID-II. *Journal of Personality Disorders*, **17**, 173–87.
36. Westen, D., Shedler, J., and Bradley, R. (2006). A prototype approach to personality disorder diagnosis, *American Journal of Psychiatry*, **163**, 846–56.
37. Millon, T., Millon, C., Davs R., *et al.* (2006). *MCMI-III manual* (3rd edn). Dicandrien Inc., Minneapolis, MN.
38. de Girolamo, G. and Reich, J.H. (1993). *Personality disorders*. WHO, Geneva.
39. Reich, J.H. (1989). Update on instruments to measure DSM-III and DSM-III-R personality disorders. *Journal of Nervous and Mental Disease*, **177**, 366–70.
40. Tyrer, P., Casey, P., and Gall, J. (1983). Relationship between neurosis and personality disorder. *British Journal of Psychiatry*, **142**, 404–8.
41. Loranger, A.W., Sartorius, N., Andreoli, A., *et al.* (1994). The international personality disorder examination: the World Health Organization/alcohol, drug abuse, and mental health administration international pilot study of personality disorders. *Archives of General Psychiatry*, **51**, 215–24.
42. Ouimette, P.C. and Klein, D.N. (1995). Test-retest stability, mood-state dependence and informant-subject concordance in the SCID-Axis II questionnaire in a non clinical sample. *Journal of Personality Disorders*, **9**, 105–11.
43. Zimmerman, M. (1994). Diagnosing personality disorders. A review of issues and research methods. *Archives of General Psychiatry*, **51**, 225–45.
44. Jackson, H.J., Gazis, J., Rudd, R.P., *et al.* (1991). Concordance between two personality disorder instruments with psychiatric inpatients. *Comprehensive Psychiatry*, **32**, 252–60.
45. Clark, L.A., Livesley, W.J., and Morey, L. (1997). Special feature: personality disorder assessment: the challenge of construct validity. *Journal of Personality Disorders*, **11**, 205–31.
46. Blashfield, R.K. and Herkov, M.J. (1996). Investigating clinician adherence to diagnosis by criteria: a replication of Morey and Ochoa (1989). *Journal of Personality Disorder*, **10**, 219–28.
47. Perry, J.C. (1992). Problems and considerations in the valid assessment of personality disorders. *American Journal of Psychiatry*, **149**, 1645–53.
48. Dodwell, D. (1988). Comparison of self-ratings within informant-ratings of pre-morbid personality on two personality rating scales. *Psychological Medicine*, **18**, 495–502.
49. Ready, R.E., Watson, D., and Clark, L.A. (2002). Psychiatric patient- and informant-reported personality. *Assessment*, **9**, 361–72.
50. Bornstein, R.F. (1998). Reconceptualizing personality disorder diagnosis in the DSM-IV: the discriminant validity challenge. *Clinical Psychology: Science and Practice*, **5**, 333–43.
51. Widiger, T.A. and Rogers, J.H. (1989). Prevalence and comorbidity of personality disorders. *Psychiatric Annals*, **19**, 132–6.

4.12.3 Specific types of personality disorder

José Luis Carrasco and Dusica Lecic-Tosevski

Cluster A personality disorders

Paranoid personality disorder

Pervasive suspiciousness, mistrust, hypersensitivity to criticism, and hostility are the essential features of paranoid personality disorder. These individuals live a restricted emotional and interpersonal life because they fear the malevolent intent of others. As a rule, paranoid people are ready to counter-attack, provoking repeated confrontations. In this way, they induce hostility and resentment in others.

The term paranoia may lead to some confusion if it is not properly delimited. Paranoid had been used as an adjective to label various delusional representations or syndromes. Kraepelin[1] differentiated paranoia as a distinct condition characterized by chronic and highly systematized delusional ideas (see Chapter 4.4). Schneider[2] described people with this paranoid personality as fanatic psychopaths, stressing their intensity, and rigidity in confrontation with others. He denied any relationship with paranoia. Freud[3] and other psychoanalysts construed the paranoid character as a pattern of mistrust and feeling of being attacked, based on distortions and externalization of the person's inner world.

Paranoid personality disorder was included in DSM-III with criteria of suspiciousness, mistrust, hypersensitivity, and restricted affectivity. This last criterion does not appear in DSM-IV and ICD-10, since restricted affectivity is neither necessary nor specific for paranoid personalities. Instead, emphasis is placed on mistrust and sensitivity to setbacks. The DSM-IV criteria for paranoid personality disorder are shown in Table 4.12.3.1.

Table 4.12.3.1 DSM-IV diagnostic criteria for paranoid personality disorder

A. A pervasive distrust and suspiciousness of others such that their motives are interpreted as malevolent, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following
1 Suspects, without sufficient basis, that others are exploiting, harming, or deceiving him or her
2 Is preoccupied with unjustified doubts about the loyalty or trustworthiness of friends or associates
3 Is reluctant to confide in others because of unwarranted fear that the information will be used maliciously against him or her
4 Reads hidden demeaning or threatening meanings into benign remarks or events
5 Persistently bears grudges, i.e. is unforgiving of insults, injuries, or slights
6 Perceives attacks on his or her character or reputation that are not apparent to others and is quick to react angrily or to counterattack
7 Has recurrent suspicions, without justification, regarding fidelity of spouse or sexual partner
B. Does not occur exclusively during the course of Schizophrenia, a Mood Disorder With Psychotic Features, or another Psychotic Disorder and is not due to the direct physiological effects of a general medical condition.

Note If criteria are met prior to the onset of Schizophrenia, add 'Premorbid', e.g. 'Paranoid Personality Disorder (Premorbid)'

(a) Epidemiology

The prevalence of paranoid personality disorder is estimated at about 0.5 to 1 per cent in the general population and at 10 to 20 per cent among psychiatric patients. The disorder is more commonly diagnosed in males.

(b) Aetiology

This personality disorder has a familial relationship with delusional disorders and with schizophrenia,[4] and has been included in the so-called schizophrenic spectrum.[5] Deficits in cortical dopamine activity may be associated with a poor conceptual organization that could in turn be responsible for suspiciousness and distorted interpretations.[6]

Mistrust and lack of confidence may reflect deficits arising in early developmental stages and resulting in a lack of basic self-confidence.[7] Lack of protective care and affective support in childhood could perhaps facilitate the development of paranoid features.

(c) Clinical picture

Paranoid individuals do not often ask for help from psychiatrists. They have no wish to be cured; instead, they believe that they have to be protected from other people's hatred and attacks. Subjects with this personality disorder suspect that others are acting to harm, exploit, or deceive them. These suspections are based not on objective evidence, but on internal conviction and an attempt to find a rational explanation for the supposed wrongs.

Paranoids are reluctant to confide in others; they tend to feel that others are plotting against them, and that the enemy may be found in unexpected places. They do not readily tell others about their suspicions. The disorder may be manifested by irritability, unusual defensive or self-protective behaviours (e.g. locking doors and closing windows and curtains to avoid being spied on, and hiding papers or documents), or emotional detachment.

Paranoid people lack confidence in others. They doubt the loyalty or trustworthiness of friends and partners, and check their behaviour repeatedly for evidence of malevolent intentions. They assume that others are not trustworthy, to the extent that they cannot believe it when friends demonstrate their loyalty. They withhold personal or significant information from friends, fearing that it will be used maliciously against them. They do not form close friendships and are often isolated. When in trouble, paranoids do not expect help from friends or others close to them; instead, they expect to be attacked or ignored.

Many of the suspicious and distrustful attitudes of paranoids are perpetuated by their intense interpersonal sensitivity. They react intensely to any comment or event that may relate to them. Hidden meanings that are demeaning and threatening may be read into benign events or the remarks of others. Unintended errors by colleagues or public servants are taken as deliberate attempts to harm or deceive them. Humorous remarks or jokes may be interpreted as attacks on their character. Paranoids are easily hurt, and their pride is easily damaged by minor critical comments or questioning. They are excessively preoccupied with attacks on their reputation or character, and minor slights may arouse major hostility and a

counter-attack. They bear grudges and harbour hostile feelings for a long time, and are unwilling to forgive the insults, injuries, or slights that they think they have received.[8]

Pathological jealousy is a common presentation of paranoid individuals. They have unreasonable doubts about the loyalty and faithfulness of their partners, based on little or no evidence. They may try to gather trivial and circumstantial facts to justify their beliefs. To avoid betrayal they attempt to gain complete control of intimate relationships, continuously questioning, and challenging partners about their whereabouts and intentions.

The interpersonal world of paranoids is a consequence of their suspiciousness and distrust. They have difficulty in relating to others, especially with close relationships. Hostility is always present and can be manifested as excessive argumentativeness, recurrent complaint and confrontation, or hostile aloofness.[8] Although they may appear rational, unemotional, and cold, the affect of paranoids is labile and oversensitive and they may be hostile, stubborn, and sarcastic. This mixture of secretive, cold, hostile, and sarcastic behaviours often elicits a hostile response in others, which confirms the paranoid person's beliefs.

Paranoids blame others for their shortcomings. They are querulous and quick to counter-attack, so that they may become involved in frequent litigation. Since they do not confide in others, paranoids need self-confidence and a sense of autonomy and independence. They need to control people who might be harmful. While they do not accept criticism, they are highly critical.

One group of paranoids are close to Schneider's 'fanatics'.[2] They have hidden grandiose fantasies of power and negative views of other people, especially those belonging to another group who come to be considered as natural enemies. They simplify issues and avoid any ambiguous perspective. Some form cults or other tightly knit groups with people who share their paranoid belief systems.

(d) Course

Paranoid features may be present in childhood and early adolescence in the form of hypersensitivity, social anxiety, poor peer relationships, and eccentricity. These features sometimes elicit teasing from other children, which in turn may aggravate the paranoid attitudes.

In situations of stress, individuals with paranoid personality disorder may respond with brief psychotic episodes. During these episodes, they may have frank delusional ideas or distorted perceptions. Some paranoid personality disorders are the premorbid state for a delusional disorder or even schizophrenia.

Individuals with this personality disorder may be at increased risk for agoraphobia, obsessive–compulsive disorder, and substance abuse or dependence. This personality disorder is often co-diagnosed with schizoid, schizotypal, narcissistic, and avoidant personality disorders.

(e) Differential diagnosis

Paranoid personality disorder should be distinguished from suspicious attitudes towards examination among immigrants, ethnic groups, or political groups. Members of these groups may display defensive and mistrustful behaviours owing to lack of familiarity with the language or the rules of a society, or in response to perceived neglect or rejection. Their behaviour may elicit further rejection from the majority, thus reinforcing the defensive behaviours.

Paranoid personality disorder is distinguished from delusional disorder, paranoid schizophrenia, and depression with psychotic symptoms, all of which are characterized by periods of persistent psychotic symptoms. Paranoid personality disorder present before the occurrence of these syndromes should be diagnosed as 'premorbid'.

People with schizotypal personality disorder are suspicious, have paranoid ideas, and keep their distance from others. However, they also experience perceptual distortions and magical thinking, and are usually odd and eccentric. Schizoid personality disorder is characterized by aloofness, coldness, and eccentricity, but these individuals usually lack prominent suspiciousness or paranoid ideation. Individuals with **avoidant personality disorder** are hypersensitive and do not confide in others. However, their lack of confidence is based on fear of being embarrassed or found inadequate rather than fear of other people's malicious intentions. Some antisocial behaviour by paranoid individuals originates in a wish for revenge or counter-attack, rather a desire for personal gain as in antisocial personality disorder. Paranoid features are often present in narcissistic individuals who fear that their imperfections could be revealed. The differential diagnosis should be based on the predominance of persistent need of praise versus persistent suspiciousness and distrust.

(f) Treatment

Antidepressant and anxiolytic treatment may be useful for anxiety and depression resulting from a paranoid response to stressful situations. Low-dose antipsychotics may be indicated during brief psychotic episodes or when ideas of reference are present.

Psychological treatment is difficult owing to the lack of insight. The approach is to attempt to gain the patient's confidence, avoiding early confrontation of distorted ideas, followed by a slow gentle attempt at cognitive restructuring.

Schizoid personality disorder

Schizoid personality disorder is characterized by a persistent pattern of social withdrawal. Schizoid individuals show discomfort in social interactions and are introverted. They are seen by others as eccentric, isolated, or lonely. DSM-IV diagnostic criteria are shown in Table 4.12.3.2.

This type of personality became recognized in the first two decades of the twentieth century. August Block's description of the shut-in personality and Eugen Bleuler's description of autism distinguished between shy and lonely persons and those who engage in relationships only in fantasy. Psychoanalysts included this term in their writings and developed an approach based on deficient object relations and individuation.[9] Some schizoid personalities have probably been sweet children who were very easy to care for, although giving less joy to their parents and eliciting less stimulation and fewer expressions of emotion than more expressive children.[7]

(a) Epidemiology

The epidemiology of schizoid personality disorder is not clearly established. Recent studies give a median prevalence of 0.5 to 1 per cent (see Chapter 4.12.5).

(b) Aetiology

A familial association may exist between schizotypal personality disorder and schizophrenia.

Table 4.12.3.2 DSM-IV diagnostic criteria for schizoid personality disorder

A.	A pervasive pattern of detachment from social relationships and a restricted range of expression of emotions in interpersonal settings, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following
	1 Neither desires nor enjoys close relationships, including being part of a family
	2 Almost always chooses solitary activities
	3 Has little, if any, interest in having sexual experiences with another person
	4 Takes pleasure in few, if any, activities
	5 Lacks close friends or confidants other than first-degree relatives
	6 Appears indifferent to the praise or criticism of others
	7 Shows emotional coldness, detachment, or flattened affectivity
B.	Does not occur exclusively during the course of Schizophrenia, a Mood Disorder With Psychotic Features, another Psychotic Disorder, or a Pervasive Developmental Disorder and is not due to the direct physiological effects of a general medical condition

Note If criteria are met prior to the onset of Schizophrenia, add 'Premorbid.' e.g. 'Schizoid Personality Disorder (Premorbid)'.

(c) Clinical picture

People with schizoid personality disorder appear cold, reserved, distant, and unsociable. They lack involvement in everyday events and in the concerns of others. They rarely tolerate eye contact, usually give short answers, and appear uneasy when asked about emotions or feelings. However, they may invest much energy in abstract ideas such as those of mathematics or philosophy.

There is a characteristic lack of emotional expression and low energy. Speech is typically slow and monotonous, and seems to lack associated emotion. Affect is excessively serious or constrained, although some inner fear may be detected by an experienced clinician. If they try to be humorous, they usually give a child-like impression. Psychomotor activity tends to be lethargic, lacking gesture, and rhythmic movement. They may seem absorbed in insignificant matters, keeping quiet and not annoying anybody, as if in their own world. They do not express joy, anger, sadness, or other emotions. Interpersonal communication tends to be formal and impersonal, although not irrational. Threats, real or imagined, are dealt with by fantasized omnipotence or resignation. Aggressive acts are infrequent.

People with schizoid personality disorder characteristically seem to lack interest in the lives and concerns of others. When in a group, they stay unnoticed and detached, seeming indifferent to criticism or praise or to the reactions of others. Schizoids are attracted to solitary hobbies, and may be successful in lonely jobs that others find difficult to tolerate. Many prefer working at night. Usually, they do not seem to suffer because of this detachment and they have no desire for closeness or intimacy. They seldom have close friends or relationships, except with immediate relatives. Their sexual lives may be poor or exist only in fantasy, and some postpone mature sexuality indefinitely. They do not usually marry, although some, especially schizoid women, may passively agree to marriage. However, schizoid individuals may make emotional attachments with animals or inanimate objects.

Schizoid personalities lack insight, and generally have a poorly developed sense of identity and a poor capacity for evaluating interpersonal events. They may appear to be self-absorbed and engage in excessive daydreaming. However, some schizoid individuals have original and creative ideas.

(d) Differential diagnosis

Schizoids have better occupational functioning than patients with **schizophrenia** or **schizotypal personality disorder**, and, although isolated, can have successful careers. Schizophrenic patients exhibit delusional thinking or hallucinations and psychotic episodes. Schizotypal individuals show greater eccentricity and oddness than schizoids, and also have perceptual and thought disturbances including magical thinking.

People with **paranoid personality disorder** may also show social detachment and lack close relationships. However, they show more social engagement than schizoids and may have a history of aggressive behaviour.

Emotional constraint is also present in **obsessive–compulsive personality disorder**, but obsessional patients are more involved in everyday life and concerns, and may be worried by criticism. People with **avoidant personality disorder** are also detached and aloof. However, although they actively avoid interpersonal contact because of fear of rejection or being found inadequate, they have an intense desire for close relationships.

(e) Course

Schizoid personality disorder is usually apparent in early childhood. As with all personality disorders, it is usually long-lasting; however, it is not necessarily lifelong although there is seldom any rapid or profound change. If their deficits are moderate and social circumstances are favourable, some schizoids achieve social and vocational adaptation.

Although this personality disorder is sometimes a precursor of schizophrenia, the number of schizoid patients who go on to develop schizophrenia is unknown.

(f) Treatment

Because they lack insight and have little motivation for change, schizoids seldom seek treatment. Motivation for change may depend on life circumstances and external pressures.

Low-dose antipsychotic medication is useful in some situations. Antidepressants and psychostimulants have also been used with some positive effects.

The psychotherapy of patients with schizoid personality disorder must be based on gaining a therapeutic alliance. Unlike paranoid patients, they may become involved in therapy and reveal fantasies, imaginary friends, and fears of unbearable dependency. Ambivalence may appear because of fear of dependence on the therapist, who must keep the necessary distance to allow a tolerable relationship for the patient.

Social skills training is sometimes useful in improving their awareness of social cues.

Schizotypal personality disorder

Schizotypia is a controversial term in psychiatry. The term was used by Kretschmer[10] to denominate the phenotypic characters that antedated the development of schizophrenia. Nevertheless, the term schizotypal personality disorder was not included in psychiatric classifications until the publication of DSM-IIIR in 1987.[11] Before

Table 4.12.3.3 DV-IV diagnostic criteria for schizotypal personality disorder

A. A pervasive pattern of social and interpersonal deficits marked by acute discomfort with, and reduced capacity for, close relationships as well as by cognitive or perceptual distortions and eccentricities of behaviour beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following
   1 Ideas of reference (excluding delusions of reference)
   2 Odd beliefs or magical thinking that influences behaviour and is inconsistent with subcultural norms (e.g. superstitiousness, belief in clairvoyance, telepathy, or 'sixth sense'; in children and adolescents, bizarre fantasies or preoccupations)
   3 Unusual perceptual experiences, including bodily illusions
   4 Odd thinking and speech (e.g. vague, circumstantial, metaphorical. overelaborate, or stereotyped)
   5 Suspiciousness or paranoid ideation
   6 Inappropriate or constricted affect
   7 Behaviour or appearance that is odd, eccentric, or peculiar
   8 Lack of close friends or confidants other than first-degree relatives
   9 Excessive social anxiety that does not diminish with familiarity and tends to be associated with paranoid fears rather than negative judgements about self

B. Does not occur exclusively during the course of Schizophrenia, a Mood Disorder With Psychotic Features, another Psychotic Disorder, or a Pervasive Developmental Disorder.

Note If criteria are met prior to the onset of Schizophrenia, add 'Premorbid', e.g. 'Schizotypal Personality Disorder (Premorbid)'.

that date, schizotypal individuals were allocated either with schizoids or with schizophrenics, and were usually labelled as latent schizophrenics or pseudoneurotic schizophrenics. However, the validity of this nosological entity is still controversial and, despite its acceptance in DSM-IV, ICD-10 does not recognize it as a separate personality disorder. Instead, ICD-10 includes the schizotypal syndrome among the psychotic disorders and not as a personality disorder, based on the biological affinities of schizotypal individuals with other schizophrenic patients. DSM-IV diagnostic criteria are shown in Table 4.12.3.3.

(a) Epidemiology

Schizotypal personality disorder is present in 0.5 to 3 per cent of the general population, with no demonstrated differences between sexes. It is more commonly diagnosed in relatives of schizophrenic patients, and the incidence is much higher in monozygotic than in dizygotic twins (33 per cent versus 4 per cent).[4]

(b) Clinical picture

The essential feature of schizotypal individuals is a pattern of peculiarity and oddness in interpersonal relationships with resulting social detachment and lack of close relationships. Because of their distorted reality processing schizotypal individuals feel intensely uncomfortable in the presence of others. Conversely, others feel uneasy in the presence of schizotypals because of their unusual ways of thinking and expressing emotions.

Like schizoids, schizotypals have a decreased desire for intimate contacts, although they may sometimes express unhappiness about their lack of relationships. As a consequence they do not have close friends or confidants other than relatives. They experience intense anxiety in social situations with unfamiliar people. They can interact if necessary, but they prefer to keep aloof because they feel different and are not interested in the concerns of others. Their anxiety in these situations is not based on feelings of inadequacy or fear of humiliation. Rather, it is due to suspicion of the motivation of others, and therefore it is not alleviated as time passes and the situation becomes more familiar. Thus schizotypals feel progressively worse and more reluctant to confide in other people.

Individuals with schizotypal personality disorder often have ideas of reference that is interpretations of casual events as having specific and unusual meanings related to themselves. However, these ideas do not achieve the pathological conviction of delusions. Similarly, these individuals may be preoccupied with superstitions or paranormal phenomena. They may feel that they may read other people's thoughts or influence their behaviour by the power of thought. Their magical thinking is often manifested by ritualized behaviours aimed at avoiding harmful events.

Perceptual disturbances are frequent in schizotypal personality disorder. An experience of a sixth sense is typical, with the 'ability' to notice someone's presence. Distorted perceptions are present in the form of sounds perceived as calling voices or shadows transformed into figures and faces.

Thought processing and speech are characteristically affected. Speech may be constructed in an unusual and idiosyncratic way-generally loose, digressive, or vague, but without actual derailment or incoherence. Responses may be either excessively concrete or far too abstract, and words may be used in unusual ways.

The interpersonal relationships of schizotypal individuals are primarily affected by paranoid and suspicious ideation. They may believe that colleagues at work want to damage their reputation. In addition to the social anxiety of these individuals, this leads to a stiff and constricted contact and affect. They are considered odd and eccentric by others: they have peculiar mannerisms, dress in an unusual and unkempt manner, adopt extravagant postures and clothing combinations, do not obey normal social conventions, and generally avoid eye contact.

(b) Course

Schizotypal features may be present in childhood and adolescence in the form of solitariness, academic underachievement, hypersensitivity, and bizarre fantasies. Schizotypals do not seek treatment because of their personality disorder, but rather because of the presentation of associated depression, dysphoria, and anxiety. In response to stressful situations, these patients may experience transient psychotic episodes lasting from minutes to hours. In some cases, clinical symptoms and duration reach the degree of brief psychotic disorder, schizophreniform disorder, or schizophrenia, with the schizotypal personality disorder as the premorbid state. The prevalence of major depressive episodes is notoriously high, as is co-diagnosis with paranoid, schizoid, avoidant, and borderline personality disorders.

(c) Differential diagnosis

Delusional disorder, **schizophrenia**, and **mood disorder with psychotic symptoms** have to be excluded based on the greater intensity and persistence of psychotic symptoms.

In childhood, it can be difficult to distinguish schizotypal personality disorder from other forms of disorders characterized by odd behaviour, isolation, eccentricity, and peculiarities of

language. These include **autistic disorder**, **Asperger's disorder**, and some **language disorders**. The differentiation with communication disorders is based on the prominence of language symptoms in these children and the compensatory efforts to communicate by gesture and other means. Autism and Asperger's disorder present an even more intense social isolation and indifference, stereotyped behaviours and interests.

Paranoid and **schizoid personality disorders** lack the perceptual and speech impairment of schizotypal personality disorder, as well as the marked eccentricity and oddness. **Avoidant personality disorder**, while including social anxiety and isolation, differs from schizotypal personality disorder in that avoidants have an intense desire for closeness, which is constrained by fear of rejection. Schizotypals do not have a desire for relationships. **Borderline personality disorder** has a high rate of co-occurrence with schizotypal personality disorder and frequently the two disorders cannot be differentiated. Brief psychotic episodes in people with borderline personality disorder are more dissociative-like and generally follow affective shifts in response to stress or frustration. Social isolation in borderline personality patients is generally due to repeated interpersonal failures rather than a persistent lack of desire for relationships and intimacy.

Finally, schizotypal personality disorder must be diagnosed in the cultural context of the patient. It should be noted that some perceptual peculiarities and magical beliefs may be due to culturally determined characteristics. For example, mind reading, voodoo, shamanism, evil eye, and so on should not be considered as personality disorders in some cultural areas.

(d) Treatment

Low-dose antipsychotic medication may be useful for ideas of reference, perceptual disturbances, and other psychotic-like symptoms. Antidepressants are effective when depressive states are associated.

The psychological management of schizotypals should include a prolonged period of gaining the confidence of the patient. However, a particularly careful approach must be adopted owing to the peculiar thought processing of these patients.

Cluster B personality disorders

Antisocial personality disorder

Antisocial personality disorder is characterized by a pattern of disregard for the safety and the rights of others, without feeling remorse. Individuals with this disorder are unreliable, manipulative, incapable of lasting relationships, and unable to conform to social norms. The disorder starts early (before the age of 15), is pervasive, and manifests in variety of contexts. Although social deviance is one of the core features of antisocial personality disorder, it is not synonymous with criminality. Antisocial personality disorder uncomplicated by other disorders is not often met in clinical settings, except forensic psychiatry. However, owing to its impact on family and social environment, it has major public health significance and has been extensively studied in academic psychiatry, psychoanalysis, law, sociology, theology, and literature.

The description of antisocial personality in the last 1970s was mainly based on criminal behaviour[12] and the disorder was conceptualized as synonymous of criminality. Later classifications modified this approach and focused on the personality traits and emotional patterns described in classic descriptions included the classic personality traits leading to the current DSM-IV and ICD-10 classification criteria for antisocial personality disorder. (Tables 4.12.3.4 and 4.12.3.5).

Table 4.12.3.4 SM-IV diagnostic criteria for antisocial personality disorder

A. There is a pervasive pattern of disregard for and violation of the rights of others occurring since age 15 years, as indicated by three (or more) of the following 1 Failure to conform to social norms with respect to lawful behaviours as indicated by repeatedly performing acts that are ground for arrest 2 Deceitfulness, as indicated by repeated lying, use of aliases, or conning others for personal profit or pleasure 3 Impulsivity or failure to plan ahead 4 Irritability and aggressiveness, as indicated by repeated physical fights or assaults 5 Reckless disregard for safety of self or others 6 Consistent irresponsibility, as indicated by repeated failure to sustain consistent work behaviour or honour financial obligations 7 Lack of remorse, as indicated by being indifferent to or rationalizing having hurt, mistreated, or stolen from another
B. The individual is at least age 18 years
C. There is evidence of conduct disorder with onset before 15 years
D. The occurrence of antisocial behaviour is not exclusively during the course of Schizophrenia or a Manic Episode.

(a) Epidemiology

A prevalence rate of about 3 per cent is consistently found in the general population, and it is more frequent in males than females, with sex ratios ranging from 2:1 to 7:1. It is more common among younger adults, people living in urban areas and lower socio-economic groups.[13]

Table 4.12.3.5 ICD-10 diagnostic criteria for disocial personality disorder

Personality disorder, usually coming to attention because of a gross disparity between behaviour and the prevailing social norms, and characterized by (a) callous unconcern for the feelings of others (b) gross and persistent attitude of irresponsibility and disregard for social norms, rules and obligations (c) incapacity to maintain enduring relationships, though having no difficulty in establishing them (d) very low tolerance to frustration and a low threshold for discharge of aggression, including violence (e) incapacity to experience guilt and to profit from experience, particularly punishment (f) marked proneness to blame others, or to offer plausible rationalizations, for the behaviour that has brought the patient into conflict with society
There may also be persistent irritability as an associated feature. Conduct disorder during childhood and adolescence, though not invariably present, may further support the diagnosis.
Includes: amoral, antisocial, psychopathic, and sociopathic personality (disorder)
Excludes: conduct disorders, emotionally unstable personality disorder

(b) Aetiology

The aetiology of antisocial personality disorder is complex and multifactorial, involving biological, early developmental, and social determinants.

Twin, adoption, and family studies have demonstrated that genetic factors strongly contribute to the development of antisocial personality.[14] Antisocial personality in males is often associated with hysteria in women of the same family which suggests that the two conditions might be alternative expressions of the same genetic endowment, belonging to 'spectrum conditions'.[15] Longitudinal studies of hyperactive children have reported high rates of later (adult) antisocial behaviour, and have suggested a 'developmental' relationship between antisocial behaviour and childhood hyperactivity.

Aggression in antisocial personality disorder is associated with indexes of reduced brain serotonin activity such as low levels of the serotonin metabolite 5-hydroxyindole-acetic acid in the cerebrospinal fluid and[16,17] low platelet monoamine oxidase activity. Reports on minimal brain dysfunctions resulting on frontal-lobe deficiencies and lack of inhibition have also been described.

Parental deprivation, inconsistent maternal care, family violence, and severe childhood physical) abuse have been reported as strong predictors for development of antisocial personality disorders.[12,18]

Social disintegration and chronic criminality can cause episodic antisocial behaviour, reflecting a normal adaptation to an abnormal social environment.[19] However, the multifactorial origin of the antisocial personality disorder and its early onset and manifestations indicate that it cannot be attributed to cultural conflicts and social determinants.

(c) Clinical features and diagnosis

Patients with antisocial personality disorder often appear quite normal, charming, and understanding. However, their history reveals disturbed functioning in the domains of behaviour and self-concept, love and sexuality, interpersonal relations, and cognitive style.[20]

Reckless behaviour unaffected by punishment is typical of antisocial individuals, who are also exploitative, manipulative, demanding, and lacking in a sense of responsibility. An easy-going hedonistic attitude may be interrupted by rage, cruelty, and violence. The absence of internalized moral values is manifested by lying, truancy, running away from home, thefts, fights, substance abuse, and illegal activities may be typical experiences, beginning in early childhood.

An impaired control of impulses and a reduced ability to anticipate the negative consequences of behaviour is typical associated to a marked intolerance to anxiety. Antisocial individuals are egocentric, and unable to feel genuine guilt and remorse. They exhibit intense and persistent anger usually expressed as hostility towards others and they have an incapacity for reflective mourning or sadness. Frequent suicide threats and attempts are also common, as is somatic preoccupation.

Interpersonal relationships of antisocial subjects are characterized by manipulation, exploitation, instability and incapacity for love, and comprehension. Sexual perversions, abuse, and paedophilia are frequent. They display deficient parenting and social dysfunction, and resistance to authorities is pronounced.

The cognitive style of antisocial subjects is characterized by glibness, superficiality of knowledge, and paranoid view of reality.

(d) Comorbidity and differential diagnosis

Antisocial personality disorder is frequently comorbid with **depression**, which usually has atypical features. **Bipolar disorder** (manic phase) and **mental retardation** (learning difficulties) should be excluded. Substance abuse may be comorbid from childhood, and antisocial behaviour may be secondary to premorbid alcoholism type 2. **Atypical schizophrenic disorder** (pseudopsychopathic schizophrenia), **temporal-lobe epilepsy**, or a **limbic-lobe syndrome** should also be excluded.

The presentation of antisocial and criminal behaviour in borderline personality disorder is frequent. However, borderline behaviours are marked by intense affective instability and reactivity and may show some remorse or guilt. Unlike antisocial patients, borderline personality disorders do not lack the capacity for intimacy and emotional investment of others and do not show sadistic behaviours. Self-aggression and suicide attempts are much more prevalent among BPD than in antisocial personality.

Aggressive and defiant behaviours are often present in histrionic personality disorder. Although some aetiologic relationship among both disorders might be possible, as described above, histrionic patients are more impulsive and emotionally driven than antisocial and display intense emotions related with attachments and losses.

(e) Course and prognosis

Antisocial behaviour is most pronounced in early adult years, and gradually decreases with age. Professional motivation and establishing a stable couple or partnership may have beneficial effects. Maturation of the personality might also take with depression or hypochondriasis emerging when rage and aggression are abandoned. Substance abuse and promiscuity are risky behaviours for developing HIV infection.

(f) Treatment

Medication is used to deal with incapacitating symptoms, such as anxiety, rage, depression, and somatic complaints. Selective serotonin reuptake inhibitors, lithium, carbamazepine, clonazepam, and other anticonvulsants have been used to control aggressive behaviour but the effects are much less pronounced than in borderline personality disorder or intermittent explosive disorder. Psychostimulants such as methylphenidate may be useful if there is evidence of attention-deficit hyperactivity disorder. Benzodiazepines are contraindicated since they might cause behavioural disinhibition.

Efficacy of psychotherapy is very little in antisocial patients. Fear of intimacy causes difficulties in establishing a therapeutic alliance, which should be oriented to find alternative defence mechanisms to acting-out and to self-defeating behaviours. Therapeutic communities based on the principles outlined by Maxwell Jones[21] with a general social adjustment as a main task, might give positive results.

Borderline personality disorder

Borderline personality disorder (BPD) is the denomination of a syndromal picture characterized by intense affective instability and impulsivity together with an unstable sense of self-identity. It is often manifested by impulsive self-aggression and suicide attempts, substance abuse, chronic feelings of emptiness, and persistent pattern of severely unstable interpersonal relationships.

The term borderline was first used by Stern[22] in 1938 to denominate a group of syndromes placed in the border between neuroses and psychoses and included also the current label of schizotypal personality disorder and a group of disorders currently classified as psychotic disorders. Only some decades later the term borderline began to be understood as a disorder of character[23] and introduced in DSM-III as a personality disorder, after being separated from schizotypal personality disorder.

Borderline personality disorder derives but is not fully equivalent to the concept of borderline personality organization developed by Kernberg.[24] BPO is a stable permanent state based on three criteria: diffuse identity, primitive defence mechanisms (splitting, denial, and projective identification), and intact reality testing. This personality organization can be found not only in BPD but also in other severe personality disorders and Axis I conditions.

Borderline personality disorder itself can be found in association with so many Axis I and Axis II disorders that its validity as an independent diagnostic category is still weak compared with other personality disorders. Some authors have suggested that borderline personality disorder reflects rather a state of severely impaired personality function than a discrete diagnostic entity.[25] Others have suggested that BPD is an atypical variant of affective disorder and should be included in the affective disorder spectrum.[26] In the ICD-10, this disorder is named as 'emotionally unstable personality disorder', with two subtypes: impulsive and borderline. Borderline subtype is specifically linked to the presence of self-identity weakness and diffusion.

(a) Epidemiology

The number of people suffering from borderline personality disorder ranges from 1.5 to 5 per cent of general population with wide differences between studies because of lack of reliable measures. The prevalence is greater in clinical samples of patients at the outpatient clinics, ranging form 10 to 15 per cent. The disorder is more common in women than in men and is commonly initiated between 18 and 35 years old.[27]

(b) Aetiology

Several factors have been associated with a higher prevalence of borderline personality disorder, including genetic, biological, and developmental findings.[28] Family studies indicate that parents of patients with BPD have a greater incidence of mood disorders but not of schizophrenia. Additionally, there is also high family incidence of antisocial personality disorder and alcoholism.

Among the biochemical findings, those indicating a brain serotonin deficiency are the more consistent. Reduced levels of 5-hydroxyindoleacetic acid in cerebrospinal fluid and blunted prolactin response to serotonin agonists have been demonstrated in association with impulsive aggression, which is a core feature of BPD.[29] Hypothalamic-pituitary—adrenal axis dysfunctions, suggesting increased feedback inhibition, as well as increased sensitivity of some areas of the amigdala, have been reported in samples of BPD patients. Current available data suggest that BPD might be associated with abnormal emotional reactivity in the limbic areas and insufficient regulatory function at the cingulated and prefrontal areas of the brain.[30]

The role of childhood trauma in the development of borderline personality disorder could be crucial. Higher incidence of childhood traumatic experiences, both for sexual/physical abuse or for neglect, has been demonstrated in these patients.[31] Other proposed developmental factors include deficiencies in self and identity development linked to attachment failures with parental figures in the early developmental phases.[32,33]

The onset of BPD needs the interaction of predisposing factors, both biological and developmental, and environmental precipitants. BPD patients seem to be extremely sensitive to frustrations in the intimate relationships, which are commonly detected at the onset of the disorder.

(c) Clinical features and diagnosis

Impulsivity and affective instability, self-aggression, identity disturbance, and unstable/intense interpersonal relationships are the most characteristic manifestations of borderline personality disorder.

Identity weakness and diffusion explain several aspects of borderline personality disorder (Table 4.12.3.6). It is clinically manifested by contradictory character traits and sense of discontinuity of the self and feelings of emptiness.[34] Probably related with this is also the intolerance to be alone and the desperate efforts to avoid abandonment by significant others. The chronic feeling of emptiness is recurrently intensified and unbearable leading to drug abuse and self-defeating behaviours.

The affect of borderline patients is chronically dysphoric and irritated. Their unstable mood is a mixture of depressed affect, anger, loneliness, and emptiness. Impulsive–aggressive behaviour is a core feature of borderline personality disorder and is related with this abnormal affective state.

Cognitive style of borderline patients are easily suggestible and frequently change their decisions. Things and people are seen in black-and-white terms. Transient and brief psychotic episodes are frequent in BPD patients associated with unstructured stressful

Table 4.12.3.6 DSM-IV diagnostic criteria for borderline personality disorder

A pervasive pattern of instability of interpersonal relationships, self-image, and affects and marked impulsivity beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following

1. Frantic efforts to avoid real or imagined abandonment. **Note:** Do not include suicidal or self-mutilating behaviour covered in Criterion 5
2. A pattern of unstable and intense interpersonal relationships characterized by alternating between extremes of idealization and devaluation
3. Identity disturbance: markedly and persistently unstable self-image or sense of self
4. Impulsivity in at least two areas that are potentially self-damaging (e.g spending, sex, substance abuse, reckless driving, binge eating). **Note:** Do not include suicidal or self-mutilating behaviour covered in Criterion 5
5. Recurrent suicidal behaviour, gestures, or threats, or self-mutilating behaviour
6. Affective instability due to a marked reactivity of mood (e.g. intense episodic dysphoria, irritability, or anxiety usually lasting a few hours and only rarely more than a few days)
7. Chronic feelings of emptiness
8. Inappropriate intense anger or difficulty controlling anger (e.g. frequent displays of temper, constant anger, recurrent physical fights)
9. Transient stress-related paranoid ideation or severe dissociative symptoms

situations. Psychotic symptoms may have a typical dissociative-like nature or present as transient self-referential ideation. Rejection sensitivity and suspiciousness usually colours the interpretations of behaviours of others.

Borderline patients are both intensely dependent and hostile towards significant others. Interpersonal relationships are unstable, intense, demanding, clinging, and characterized by alternation between extremes of idealization and devaluation, deriving from the defence mechanism of splitting. Infatuations are followed by devaluation of love objects. There is a tendency towards promiscuity and perversions.

(d) Comorbidity and differential diagnosis

Borderline personality disorder is frequently comorbid with **affective disorders** (major depression, dysthymia, and 'double depression'), **anxiety disorders**, **somatization disorder**, **post-traumatic stress disorder**, and **alcohol abuse**.

Differential diagnosis has to be made with **type II bipolar disorder**. Bipolar patients more often present emotional lability from sadness/apathy to euphoria while BPD patients are characterized by intense and reactive affective instability and shift rather from sadness to tolerable dysphoria. Intermittent explosive disorder also shows impulsive and aggressive behaviours but lacks identity disturbances and affective instability typical of BPD. Mild mental retardation might have intense irritability, lability, and impulsive/aggressive behaviour but lacks chronic feelings of emptiness and self-identity diffusion. Transient psychotic episodes and stress-related referential ideas of BPD should be differentiated from pervasive psychotic-like experiences of schizotypal personality disorder.

Borderline personality disorder has been shown to be associated with most personality disorders, especially with those from the dramatic cluster. The high prevalence of comorbid personality disorders may result from overlapping of diagnostic criteria or reflect the confirmation that there is the underlying borderline personality organization of all severe personality disorders. However, some features like chronic feelings of emptiness, self-mutilation, short-lived psychotic episodes, intense and episodic drug abuse, and intense ambivalent dependency in close relationships suggest a primary diagnosis of BPD.

(e) Course and outcome

Borderline patients often experience profound dysfunction in many important aspects of life including education, jobs, partner relationships, and marriage. Alcohol and psychosexual problems are also frequent. Repeated suicide attempts and premature death from suicide are frequent complications of borderline personality disorder; therefore suicidal gestures and intentions should be always taken seriously. It has been reported that 8 to 10 per cent per cent of all persons with borderline personality disorder commit suicide.[35]

The long-term outcome of borderline patients has not been studied, but the diagnosis is rarely made in patients aged over 40. It is speculated that neural structures and defence mechanisms mature with age and that these changes, together with social learning, reduce symptomatology.

(f) Treatment

Pharmacotherapy is targeted to symptoms such as affective changes (depression, anxiety, rage, dysphoria), cognitive disturbances (brief psychotic episodes or interpretative distortions), and impulsive behavioural dyscontrol.[36] New antidepressants, including SSRIs and venlafaxine, have shown positive effects in treating a broad spectrum of acute symptoms, including depression, hostility, irritability, anxiety, obsessive–compulsive symptoms, suicidal attempts, and impulsivity. Antipsychotics and anticonvulsants may help some patients, even in the absence of EEG or organic changes. There are still no clinical predictors for efficacy, therefore, a pragmatic approach is indicated with patients being treated with two or three drugs in a sequential trial.[36] Suicidal and abusive use of drugs prescribed and non-compliance may be serious problems for treatment of BPD.

Various psychotherapeutical modalities are used, including psychodynamic psychotherapy, supportive psychotherapy, and dialectical–behavioural therapy.

A more structured form of psychoanalytic approach, involving expressive psychotherapy and an active role of psychotherapist have been proposed specifically for BPD. The aims are confrontation of maladaptive defences and interpretation of transference, focusing on the 'here and now', without attempting the achievement of a full genetic reconstruction.[37]

Short-term psychotherapy is useful for managing crises or introducing long-term forms of therapy. Supportive psychotherapy is suggested for more fragile borderline patients, who are prone to serious regression in treatment. In practice, supportive therapy, with a psychoeducational component, has been the most frequently used form of treatment for borderline personality disorder. It is also possible to combine elements of intensive dynamic therapy with supportive therapy, depending on the ego strength of the patients.[38]

Dialectical–behavioural therapy[39] is based on cognitive techniques associated with reality confrontation. The major aim is emotional self-regulation and behavioural self-control and is particularly indicated for control of suicidal and impulsive behaviours and treatment of emotional reactivity.

Histrionic personality disorder

(a) Definition

Histrionic personality disorder derives from the concept of hysterical personality, supported by descriptive literature and clinical tradition but not so much by valid empirical research. It is characterized by excessive emotionality and attention seeking, and by dramatic, colourful, and extroverted behaviour. Egocentric, dependent, and demanding interpersonal relationships are typical of this disorder, which begins in early adulthood and is present in a variety of contexts. The DSM-IV and ICD-10 diagnostic criteria for this disorder are shown in tables 4.12.3.7 and 4.12.3.8.

It was included in scientific medicine by Kraepelin,[40] who described multiple symptoms, including capricious and inconsistent behaviour, histrionic exaggeration, and a life of illness, which captured the core pattern of the illness. Freud[41] recognized the relationship between hysterical neurosis and what he called the 'erotic personality, whose major goal in life is the desire to love or above all to be loved'. Psychoanalytic theorists often distinguish hysterical ('healthier') and histrionic ('sicker') personalities, where the latter has borderline organization and is an exaggeration of the former. Differences between the two concepts are shown in Table 4.12.3.9.

Table 4.12.3.7 DSM-IV diagnostic criteria for histrionic personality disorder

A pervasive pattern of excessive emotionality and attention seeking, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following

1 Is uncomfortable in situations in which he or she is not the centre of attention
2 Interaction with others is often characterized by inappropriate sexually seductive or provocative behaviour
3 Displays rapidly shifting and shallow expression of emotions
4 Consistently uses physical appearance to draw attention to self
5 Has a style of speech that is excessively impressionistic and lacking in detail
6 Shows self-dramatization, theatricality, and exaggerated expression of emotion
7 Is suggestible, i.e. easily influenced by others or circumstances
8 Considers relationships to be more intimate than they actually are

(b) Epidemiology

The prevalence is found to be 2 to 3 per cent of the general population and 10 to 15 per cent in clinical settings.[42] No significant difference has been found in terms of race and education and is more frequently diagnosed in women than in men. However, some cultural biases associated with sex role stereotypes and emotional expressiveness could lead the lower diagnostic rates in men.[43]

(c) Aetiology

Some studies suggest that histrionic personality runs in families, but evidence for a biological or learning transmission is not yet consistent.[44] Traits such as extraversion, emotional expression, and reward dependence have a strong genetic origin and might be constitutional. Biological findings associated with impulsivity, such as serotonin deficiency, can be found in histrionic patients with marked emotional instability and impulsive behaviours. It has been proposed that histrionic personality in women is genotypically linked to antisocial personality in men.[45]

From a development perspective, histrionic personality is considered to be a result of abnormally intense attachment with parental figures. The erotization component of this attachment in the oedipical phase was classically emphasized in the psychoanalytic research, although recent approaches suggest that there oral/dependent factors, derived from anomalies in earlier phases are of greater importance in the development of the disorder.[46]

Table 4.12.3.8 ICD-10 diagnostic criteria for histrionic personality disorder

Personality disorder characterized by at least three of the following:

(a) self-dramatization, theatricality, exaggerated expression of emotions
(b) suggestibility, easily influenced by others or by circumstances
(c) shallow and labile affectivity
(d) continually seeking for excitement, appreciation by others, and activities in which the patient is the centre of attention
(e) Inappropriate seductiveness in appearance or behaviour
(f) Overconcern with physical attractiveness

Associated features may include egocentricity, self-indulgence, continuous longing for appreciation, feelings that are easily hurt, and persistent manipulative behaviour to achieve own needs

Includes: hysterical and psycho-infantile personality (disorder)

Table 4.12.3.9 Types of histrionism

Hysterical personality	Histrionic personalitya
Neurotic personality organization	Borderline personality organization
Integrated identity	Diffuse identity
Predominance of repression	Predominance of splitting
Intact reality testing	Intact reality testing (proneness to distortion)
Integrated superego	Marked superego defects
Strongly bonded families	Disturbed, often broken families
Steady educational and vocational careers	Erratic careers
Capable of maintaining long-term friendships	Chaotic interpersonal relationships
Suggestible in triangular relationships	Diffuse suggestibility
Inauthenticity	Multiple identifications
Changing moods	Frequent dysphoria
Sexual inhibition	Promiscuity, perverse tendencies
Competitiveness with the same sex	Less differentiated behaviour toward sexes
Genital traits	Oral/pregenital traits

[a] Includes hysteroid, hysteriform, oral hysteric, and sick hysteric personality disorders. After Akhtar.[31]

(d) Clinical features and diagnosis

Emotionality, dramatization, exhibitionism, egocentricity, and sexual provocativeness are typical of histrionic personalities. However, behavioural expression is not always as manifested and other emotional and cognitive aspects may help for diagnosis.

Histrionic individuals are inappropriately seductive and aggressively demanding. They are self-centred, crave for novelty and excitement, and are prone to temper tantrums. Histrionic subjects are hyperemotional and impulsive, but their emotional enthusiasm is superficial and transient and their mood is labile. They describe their emotions in an inappropriate and exaggerated way in an attempt to obtain attention. Histrionic individuals are suggestible, demanding, accusative, and guilt inducing. In intense stressful situations they can show a dissociative-type of indifference and infantilism.

Histrionic personalities are inclined to sexualize all non-sexual relations, often indiscriminately, not only with a chosen partner but also with a wide variety of persons in various social, occupational, and professional relationships. Pseudosexuality is often accompanied by frigidity. A romantic outlook or a superficially adoring attitude often disguises needs for dependency and emotional attachment to a significant protective figure. Sicker individuals may be promiscuous, and may engage in multiple perverse activities.

Cognitive style is global, impressionistic, and diffuse, and lacks sharpness of detail. Non-verbal communication is better than verbal, speech is inhibited, and education is often superficial. Speech may show malapropisms or slips of the tongue.

The basic belief of histrionic personalities is that others should be impressed, and their basic strategy is dramatization. They blossom when they are the centre of attention and are highly disappointed when they are not, and draw attention to themselves by acting and speaking in a charming flirtatious way. Histrionic individuals quickly respond to others in an intimate way, often treating superficial acquaintances as if they were friends.

(e) Comorbidity and differential diagnosis

There is current evidence that **somatization disorder** (Briquet's syndrome) and **conversion disorder** can occur in conjunction with histrionic personality disorder, as well as **dissociative disorder** and **brief reactive psychosis**.[21] Differential diagnosis should not be difficult, because histrionic personality disorder is a lifelong disturbance with a chronic course, unlike Axis I disorders which are episodic. **Hypomanic and manic states** may be accompanied by seductive behaviour and exaggerated expression of emotions, but can be distinguished from histrionic personality by their episodic nature and the presence of other characteristic symptoms.

A great deal of overlap has been found between histrionic personality disorder and other **Axis II disorders**, defined by DSM-IIIR criteria; of these, the borderline, narcissistic, antisocial, and dependent personality disorders are the most frequent.

Borderline patients have more chaotic interpersonal relationships, make frequent suicide attempts, and are prone to regressive episodes of a psychotic nature. Histrionic individuals share sexual promiscuity, corruptibility, shallow emotions, and a self-centred attitude with antisocial personalities.[47] However, they do not show sustained, calculated, and ruthless disregard of social norms. Narcissists may also seek attention, but they want to be admired for their superiority while histrionic persons are clinging and dependent. Unlike narcissists, histrionic individuals have empathy for other persons. However, the features of the two disorders can be combined.

(f) Course and prognosis

Depressive symptoms, suicide attempts, and frequent use of medical services are common. Histrionic personality may gradually improve with age, as if a maturation of histrionic infantilism occurs over the years.

(g) Treatment

Depressive and anxious symptoms are frequent in histrionic personality disorder and can be alleviated with the use of antidepressants and anxiolytic medications. However, extreme care should be taken for treatment due to the vulnerability of these patients for medication abuse and non-compliance.

Supportive therapy is indicated for acutely distressed histrionic patients, as well as for those at the sicker end of the continuum. Psychoanalytic techniques in histrionic patients[47] are oriented to clarification of the patient's covert inner feelings. Patients are often demanding, want to take a special place in the therapist's life, and act out during therapy sessions, threatening to abandon treatment or undertake dangerous actions. Clear limits should be set and demanding dependent behaviour should not be rewarded.

Narcissistic personality disorder

Narcissistic personality disorder is characterized by an exaggerated sense of self-importance with a lack of sustained positive regard for others. Grandiosity (in fantasy or behaviour) and constant craving for admiration and external gratification are additional features of this disorder. They are present in a variety of contexts and begin by early adulthood.

(a) Historical perspective

The term narcissism originates from the Greek myth of Narcissus who was infatuated with his own reflection in the mirror-lake. Its contemporary usage has many meanings and implications, from its colloquial usage denoting self-centred persons, often with pejorative connotations, to a pathological clinical syndrome. Despite the popularity of the construct, there is still considerable disagreement on the aetiology and phenomenology of narcissistic personality disorder. There is little empirical evidence regarding its description, clinical utility, and validity.

A narcissistic personality has a pathological grandiose self, which hides a diffuse and aimless inner identity. Kernberg argues that self-hatred, rather than self-love, lies at the root of pathological narcissism, and distinguishes between narcissism in the broad sense and the specific pathological structures of the narcissistic personality. According to Kernberg, narcissistic patients function on a borderline level. Malignant narcissism,[37] which develops when primitive aggression infiltrates the pathological grandiose self, lies at the extreme end of a continuum. It is a combination of narcissistic personality disorder, antisocial behaviour, egosyntonic aggression or sadism directed against others, and a strong paranoid orientation (Table 4.12.3.10).

Narcissistic personality disorder was officially accepted in DSM-III. Somewhat refined criteria were adopted in DSM-IV (Table 4.12.3.11), because some studies showed a substantial lack of diagnostic reliability when the DSM-III criteria were used. Narcissistic personality disorder is not included in ICD-10, being mentioned only in the category 'Other specific personality disorders'.

Table 4.12.3.10 Types of narcissism

Normal	Pathological	Malignant
Infantile	Grandiose self-image	Grandiose self-image
Regression or fixation to infantile narcissistic goals in personality disorders (personality traits)	Low self-esteem Primitive defences Superego defects Borderline organization	Aggression Paranoid traits Antisocial behaviour Explosive traits Borderline organization
Adult		
Healthy self-esteem regulated by normal self-structure		
Integrated object representations		
Capacity for deep object relations		
Integrated superego		

After Kernberg.[32,37]

Table 4.12.3.11 DSM-IV diagnostic criteria for narcissistic personality disorder

A pervasive pattern of grandiosity (in fantasy or behaviour), need for admiration, and lack of empathy, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following

1 Has a grandiose sense of self-importance (e.g. exaggerates achievements and talents, expects to be recognized as superior without commensurate achievements)
2 Is preoccupied with fantasies of unlimited success, power, brilliance, beauty, or ideal love
3 Believes that he or she is 'special' and unique and can only be understood by, or should associate with, other special or high-status people (or institutions)
4 Requires excessive admiration
5 Has a sense of entitlement, i.e. unreasonable expectations of especially favourable treatment or automatic compliance with his or her expectations
6 Is interpersonally exploitative, i.e. takes advantage of others to achieve his or her own ends
7 Lacks empathy, i.e. is unwilling to recognize or identify with the feelings and needs of others
8 Is often envious of others or believes that others are envious of him or her
9 Shows arrogant haughty behaviours or attitudes

(b) Epidemiology

The prevalence of narcissistic personality disorder in the community has been found to be around 0.5 per cent.[48] Its prevalence in clinical populations is estimated to range from 1 to 3 per cent, and is more frequently diagnosed among males.

(c) Aetiology

Although no demonstrating evidence is yet available, some aspects of narcissism might be related with inappropriate seeking for excitement and reward and associated to monoamine function abnormalities at the mesolimbic reward systems.

Severe frustration with early objects is considered important in the defensive genesis of narcissistic personality disorder. Behind the compensatory grandiose self, a hungry and inferior real self-resides, as the core problem of narcissistic personality disorder. Often, high parental expectations and harsh criticism of the child is present in the family.

(d) Clinical characteristics and diagnosis

Narcissistic patients only seek for help when depressed or involved in interpersonal problems.

An often engaging and attractive appearance masks intense preoccupation with self-regard and an unusual absence of concern for others. Narcissistic individuals may be energetic, capable of consistent work, and socially successful, but this is done in order to obtain admiration. Their successes provide no inner satisfaction and always end with frustration and a feeling of emptiness. Narcissistic grandiosity is often masked by opposing tendencies (false modesty, social aloofness, and a pretended contempt for status). Pathological lying is frequent.

Narcissistic individuals feel bored when the external glitter wears off and there are no new sources to feed their self-esteem, which is extremely fragile. Lacking emotional depth, they do not have genuine feelings of sadness or longing. Anger and resentment laden with vengeful wishes are frequent as a reaction to injured self-esteem. Chronic intense envy is present, as are defences against such envy, particularly devaluation, omnipotent control, and narcissistic withdrawal. They have frequent mood swings, and hypomanic exaltation is often part of the clinical picture.

Narcissistic persons are unable to fall in love, and only have fantasies of ideal love. Sexuality is trivialized, and intercourse is a purely physical pleasure. Promiscuity, perverse fantasies, devaluation of objects, and boredom in relationships are frequent.

Interpersonal relationships are frequently manipulative and exploitative. They idealize people whom they expect to feed their narcissism, but depreciate and treat with contempt others (often former idols) from whom they do not expect to receive anything. They lack empathy and concern for others, who are welcome only as an applauding crowd and as mirrors of success.

Typical defence mechanisms (omnipotence, omniscience, intellectualization, rationalization, idealization, and devaluation) are derived from splitting.

(e) Comorbidity and differential diagnosis

Narcissistic personality disorder is often comorbid with major depression, dysthymic disorder, substance abuse, and anorexia nervosa. Patients meeting criteria for narcissistic disorder have a high overlap with **histrionic, borderline, and antisocial personality disorders**, and also with **schizotypal, paranoid, and passive–aggressive personality disorders**.

Narcissistic personality disorder may display some features of bipolar disorder (manic and hypomanic episodes). However, the mood swings are of limited duration and change rapidly, while insight is maintained and the general integrity of the personality is preserved.

Narcissistic personality disorder is strikingly similar to borderline personality disorder. Phenomenologically, grandiosity is the best discriminator between the two disorders.[49] In narcissistic personality disorder, there is also better impulse control, greater social adjustment and anxiety tolerance, less frequent suicide attempts, and less danger of regressive fragmentation and psychotic episodes.

Narcissistic individuals, especially those manifesting malignant narcissism, may demonstrate antisocial behaviour. However, antisocial individuals are more impulsive and less capable of concentrating on work and career, and they are devoid of guilt feelings. Similarities with histrionic and obsessive–compulsive personalities are superficial, since people with these disorders have a capacity for empathy and a concern and love for others.

(f) Course and prognosis

Patients often become depressed or defensively hypomanic during middle age, when their internal life gradually deteriorates owing to a vicious circle of frustrations and disappointments and diminishing narcissistic supplies. Hypochondriasis and anxiety disorders are frequent complications.

(g) Treatment

Anxiolytic agents and antidepressants may be helpful for alleviating target episodes of mood and anxiety symptoms.

(i) Psychotherapy

Individual psychotherapy is aimed to the analysis of idealizing transference and interpretation of self-grandiosity. However, during

the first stages only supportive therapy is recommended with interpretations delayed until confident and integrated attachment with therapist is achieved.

The treatment of narcissistic individuals inevitably arouses serious countertransference problems, because of the emotional detachment, demanding behaviour, and devaluative actions of narcissistic patients. The therapist should have worked through his or her own narcissism and retain an empathic and non-judgemental attitude.

Cluster C personality disorders

Avoidant personality disorder

Avoidant personality disorder was first introduced into psychiatric classification in DSM-III.[11] Before this, such patients were included among the schizoid or dependent patients. The emphasis on avoidant behaviour as a consequence of an intense sensitivity to rejection led to the differentiation of this new personality type.

The characteristic behaviour of the avoidant personality is active isolation from the social environment. Unlike schizoids, who are characterized by passive social isolation, avoidant subjects turn inward to protect themselves because they are extremely sensitive to rejection.[50] They desire interpersonal relationships but they avoid any chance of disapproval. Thus, only relationships that are likely to lead to complete non-critical acceptance are established. The extreme sensitivity to criticism is based on intense feelings of inferiority, poor self-concept and low self-esteem. This disorder is termed anxious personality disorder in ICD-10, since anxiety is considered to be the basic affective feature. The DSM-IV criteria for avoidant personality disorder are shown in Table 4.12.3.12.

(a) Epidemiology

The prevalence of avoidant personality disorder is estimated to be less than 1 per cent in the general population, but almost 10 per cent in clinical populations. No differences between sexes are found.

(b) Aetiology

Some familial transmission is possible, perhaps involving learning and identification, but genetic transmission may also be involved.[51] The biological mechanisms involved in anxiety disorders and social phobia may have a role in the development of this personality disorder. It has been suggested that hypersensitivity of brain areas involved in the separation-anxiety response and overactivity of serotonin limbic neuronal circuits may underlie the avoidant temperament trait.[51]

Psychosocial factors mediate the extent to which biological vulnerability is expressed. Children who are belittled, criticized, and rejected by parents have decreased self-esteem, resulting in social avoidance. These problems are reinforced and perpetuated at school and may generate the expectation of rejection from everyone.[50]

Table 4.12.3.12 DSM-IV diagnostic criteria for avoidant personality disorder

A pervasive pattern of social inhibition, feelings of inadequacy, and to negative evaluation, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following

1 Avoids occupational activities that involve significant interpersonal contact, because of fears of criticism, disapproval, or rejection
2 Is unwilling to get involved with people unless certain of being liked
3 Shows restraint within intimate relationships because of the fear of being shamed or ridiculed
4 Is preoccupied with being criticized or rejected in social situations
5 Is inhibited in new interpersonal situations because of feelings of inadequacy
6 Views self as socially inept, personally unappealing, or inferior to others
7 Is unusually reluctant to take personal risks or to engage in any new activities because they may prove embarrassing

(c) Clinical picture

Avoidant people are characterized by extreme shyness. They appear distant from others and do not express wishes, demands, or opinions. However, this behaviour contrasts with an extreme internal need for warmth and closeness. This contradiction is explained by an exaggerated sensitivity to rejection by others. People with this personality disorder are easily hurt and humiliated by comments from others, which they misinterpret as degrading and disapproving. They tend to be shy, quiet, and inhibited. They say nothing rather than risk being wrong, and they react strongly to any possible indications of mockery or criticism. They usually appear anxious, self-doubting, and insecure when speaking, often use self-defeating expressions, and try to please others. Their tense and fearful demeanour may elicit ridicule from others, which confirms their insecurity. They are concerned with reacting to scrutiny by blushing or crying, which is a cause of further interpersonal avoidance. These patients often choose occupations where no social interaction is needed, and strongly avoid talking in public. The avoidant person lacks intimate relationships with friends or sexual partners unless they anticipate non-critical acceptance.

Patients with avoidant personality disorder perceive themselves as inept and inadequate, and assume that they are unattractive. They tend to see others as negative and potentially harmful. They are inattentive and repeatedly distracted by intrusive thoughts, but they attend intensely to signals of rejection. These people tend to make negative evaluations of situations and exaggerate risk. They have a low tolerance for dysphoric affects, which they avoid by escaping. Escape from reality through fantasy is their usual way of satisfying their needs and relieving frustration.

At interview they may be quite open if they feel accepted. This happens when good rapport is made, which is often easier in clinical than in social situations. However, social limitation outside the office may be intense. Avoidant patients usually feel ashamed about many aspects of their lives and are excessively self-critical, although most of the concerns expressed seem to be trivial.

(d) Course

Avoidant personality disorder may follow childhood fear of strangers and shyness and isolation during school years. However, most shyness in childhood gradually dissipates in adolescence. When it evolves into avoidant personality disorder, the shyness may worsen in adolescence when social and interpersonal relationships become more complex and demanding. The disorder tends to remit or to become less evident in older people.

Avoidant personality disorder is often associated with depressive episodes, dysthymia, and anxiety disorders, particularly social phobia.

(e) Differential diagnosis

It is often difficult to differentiate avoidant personality disorder and social phobia of the generalized type. Impairment and distress due to the phobic situations is more intense in social phobia, which may have started in middle adulthood rather than adolescence. It is not clear whether the disorders are alternative manifestations of the same condition, or are separate disorders.

Hypersensitivity to rejection and criticism, low self-esteem, and feelings of inadequacy are also features of dependent personality disorder. While the avoidant patient avoids contact, the dependent patient focuses on being cared for. However, the disorders often co-occur and must be diagnosed together.

Schizoid and schizotypal personality disorder are also characterized by social isolation. However, avoidants want to have relationships and suffer for their isolation, while schizoids and schizotypals accept isolation.

People with paranoid personality disorders lack confidence in others. However, avoidants do not confide in others because they fear being found inadequate, whereas paranoids fear malicious intent.

(f) Treatment

Anxiety and hypersensitivity to rejection may improve with anxiolytic medication, β-blockers, monoamine oxidase inhibitors, and antidepressant medication. Medication should be combined with psychological treatment based on reinforcing assertiveness and self-esteem, and restructuring cognitive distortions concerning the self and others. Conscious and unconscious dependency needs should be addressed.

Dependent personality disorder (JLC)

Individuals with this disorder show a persistent and global pattern of behaviours directed at avoidance of the loss of intimate others. To attain this goal, they relinquish their own needs, opinions, expression of feelings, and even their self-identity. In exchange, they get others to take over responsibility for major areas of their lives and to protect them. Their self-concept is characterized by weakness and helplessness, while others are perceived as powerful and protective.

These people were formerly included in different classificatory categories. They belong to the abulic type of Kraepelin and of Schneider[2] and were considered as immature personalities.

(a) Aetiology

Classic hypotheses attributing dependent personality to fixation in oral phase of psychosexual development have given way to others indicating that the cause is rather deprivation than overgratification in the oral phase.

Dependent personality was recognized first in DSM-III.[11] The description has changed little in DSM-IV (Table 4.12.3.13) and ICD-10.[52] The crucial feature in both systems is the urgent need of patients to be cared for by others, with dependence, attachment, and fear of abandonment. Lack of self-confidence was required in DSM-III, but was eliminated from recent classifications because it is not specific to this disorder.

(b) Clinical picture

Dependent patients are passive. They rarely express needs or feelings, especially those that are sexual or aggressive. They tend to avoid responsibilities or decisions in major areas of their lives, such as work and financial or interpersonal relationships. Instead they get others, particularly family or partner, to decide for them or to provide continuous guidance. They depend on others (often one other person, usually the partner or a parent) to decide where they should go, what they should do, and even which clothes they should wear. They manifest self-doubt, pessimism, and a need for affection. They lack aggressiveness and appear helpless. The dependent patient avoids jobs that demand taking responsibility and managing others, and becomes anxious when forced into such situations. These patients seek intensely for companionship and do not tolerate being alone. They may function at an adequate level if in a close and protective relationship, but when left alone they are unable to survive. They believe that they are incapable of functioning independently and require constant assistance. They do not initiate projects, but wait for others who, they believe, will do them better. However, dependent individuals can perform such tasks for other people whom they want to please and to whom they want to attach themselves.

They accept unpleasant tasks, are self-sacrificing, and tolerate verbal, physical, or sexual abuse. Abusive relationships may be accepted as long as the attachment is preserved and is not excessively distorted.

An excessive and unrealistic fear of abandonment is constant in dependent individuals. When an intimate relationship is terminated by separation or death, dependent individuals urgently seek another person who will provide the care and support they seek. Thus they become rapidly and indiscriminately attached to other persons when left alone.

These people are pessimistic, self-doubting, and have low self-esteem. They belittle their capacities and successes and present themselves as inept. They take criticism as a proof of their ineptness and confirmation of their lack of self-confidence.

Table 4.12.3.13 DSM-IV diagnostic criteria for dependent personality disorder

A pervasive and excessive need to be taken care of that leads to submissive and clinging behaviour and fears of separation, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following

1. Has difficulty making everyday decisions without an excessive amount of advice and reassurance from others
2. Needs others to assume responsibility for most major areas of his or her life
3. Has difficulty expressing disagreement with others because of fear of loss of support or approval. **Note** Do not include realistic fears of retribution
4. Has difficulty initiating projects or doing things on his or her own (because of a lack of self-confidence in judgement or abilities rather than a lack of motivation or energy)
5. Goes to excessive lengths to obtain nurturance and support from others, to the point of volunteering to do things that are unpleasant
6. Feels uncomfortable or helpless when alone because of exaggerated fears of being unable to care for himself or herself
7. Urgently seeks another relationship as a source of care and support when a close relationship ends
8. Is unrealistically preoccupied with fears of being left to take care of himself or herself

(c) Epidemiology

Recent studies have found a median prevalence of 0.7 per cent (see Chapter 4.12.5). Although dependent personality disorder is diagnosed more frequently in women, structured interviews have not shown significant differences between the sexes. Cultural factors may affect the reported prevalence, as passivity, politeness, and submission are normal in some societies.

(d) Course

Dependent personality features present in adolescence may evolve positively in adulthood or lead to a personality disorder. Dependent individuals are at increased risk of depressive, anxiety, and adjustment disorders, particularly in relation to loss of close relationships. Dependent personality disorder may follow separation anxiety in childhood, or chronic physical illnesses in childhood requiring long periods of care and attention.

(e) Differential diagnosis

Dependent personality disorder has some similarities with histrionic personality disorder. Histrionic patients, like dependent patients, adjust their conduct to please other people. Their lives are centred in these others. However, people with histrionic personality disorder obtain attention and care by seductive or manipulative behaviours, whereas people with dependent personality disorder wait passively for others to care for them.[7]

Like people with avoidant personality disorder, dependent individuals may feel devastated by minor criticism or lack of attention from others. However, they lack the sense of embarrassment and social shyness of the avoidant, and fear loneliness or abandonment.

People with dependent and borderline personality disorders share an excessive fear of abandonment. However, borderline individuals react to separation with feelings of emptiness and rage, and are demanding, in contrast with the submissive and appeasing attitude of dependent individuals, which is directed towards finding another person to provide support.

Dependent personality disorder must be differentiated from normal dependent behaviours in specific life situations; for example, elderly people with chronic or debilitating disease may become dependent.

(f) Treatment

Pharmacological treatment is indicated only when depressive or anxiety symptoms are present, especially when associated with separation or loss.

In psychotherapy, the therapist must avoid the development of excessively dependent attachments. Self-confidence and self-esteem should be enhanced and the patient helped to enjoy the feeling of personal autonomy and independence. Cognitive restructuring and social skills training are often useful in bringing these changes about.

Obsessive–compulsive (anankastic) personality disorder

While DSM-IV labels this personality disorder as obsessive–compulsive personality disorder (Table 4.12.3.14), ICD-10 prefers the term anankastic, previously used in European psychiatry to refer to fearful, insecure, and compulsive individuals. The cardinal feature of this disorder is an exaggerated and pervasive attempt to control. Anankastic patients need to control those who are close to them, to control every uncertainty, and to control their own thoughts and emotions. The anankastic lacks an internal sense of security and tries to make the external world totally predictable. The anankastic is afraid of his own internal aggressive drives and avoids free emotional expression. Others perceive this kind of personality as characterized by inflexibility and stubborn inefficiency.

Table 4.12.3.14 DSM-IV diagnostic criteria for obsessive–compulsive personality disorder

A pervasive pattern of preoccupation with orderliness, perfectionism, and mental and interpersonal control, at the expense of flexibility, openness, and effciency, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following

1. Is preoccupied with details, rules, lists, order, organization, or schedules to the extent that the major point of the activity is lost
2. Shows perfectionism that interferes with task completion (e.g. is unable to complete a project because his or her own overly strict standards are not met)
3. Is excessively devoted to work and productivity to the exclusion of leisure activities and friendships (not accounted for by obvious economic necessity)
4. Is overconscientious, scrupulous, and inflexible about matters of morality, ethics, or values (not accounted for by cultural or religious identification)
5. Is unable to discard worn-out or worthless objects even when they have no sentimental value
6. Is reluctant to delegate tasks or to work with others unless they submit to exactly his or her way of doing things
7. Adopts a miserly spending style toward both self and others; money is viewed as something to be hoarded for future catastrophes
8. Shows rigidity and stubbornness

(a) Epidemiology

The prevalence of obsessive–compulsive personality disorders is about 1 per cent in community samples and up to 10 per cent in psychiatric patients, especially those with depressive and anxiety disorders. It is most frequent among males. Some obsessive–compulsive traits are sanctioned in some cultures, and a personality disorder should not be diagnosed unless the traits are markedly beyond the average for the culture.

(b) Aetiology

Biological factors and learning seem to be involved in the aetiology of obsessive–compulsive personality disorder. The personality may be partly inherited.[53] Early psychodynamic theories linked obsessive personality to the anal phase of psychosexual development between the ages of 2 and 4, when libidinal drives come into conflict with parental attempts to socialize the child, especially in sphincter control and toilet training. Later psychoanalytic theory[54] emphasized earlier manifestations of the child's autonomy versus parental wishes. The expression of drives and emotions, including anger, is shaped by parental responses and may evoke shame and criticism.

This dynamic sequence is reinforced in societies which are strongly influenced by the Protestant work ethic, in families where individual emotions are subordinated to the group, and in societies in which open expression of emotions is discouraged.

(c) Clinical picture

The behaviour of an obsessive–compulsive personality has been consistently described as one of orderliness. The patient is preoccupied with details, and pays attention to rules, procedures, schedules, and punctuality. Patients with obsessional personalities often produce their own detailed lists of symptoms and are annoyed if any item is neglected or misinterpreted. They repeat actions and check for mistakes, despite the inconvenience and annoyance that result from this behaviour. As a consequence, their conduct is frequently inefficient. For example, the combination of unproductive perfectionism and rigidity may lead to difficulty in finishing a written report on time because of excessive correction and rewriting. Since this striving for perfection and order is time consuming, other areas of their lives often appear disorganized. One room or one desk drawer may fall into disarray, or parts of their social or family lives may be disorganized.

People with obsessive–compulsive personality focus on work and productivity. It is difficult for them to take vacations or even to have free time. They do not enjoy leisure activity, which they may consider a waste of time. Often, they need to take work home to alleviate their anxiety. Hobbies and leisure pursuits become formally organized activities. They insist on perfect performance of sports or games and transform them into a serious task requiring careful organization and hard work. Leisure activities may be an unpleasant experience for the others involved, owing to the insistence on rules and standards.

Stubbornness is another characteristic of these people. They need things to be done in their way, and realistic arguments do not usually make them change their insistence. They need others to submit to their way of doing things, and often believe that no one can do the tasks as perfectly as they can. They give detailed instructions, insisting that their way is the only way of doing things, and are irritated if others suggest alternatives. Therefore, they generally insist in doing everything themselves and are unable to delegate, which increases their inefficiency at work. Paradoxically, their stubbornness is associated with doubt. Indecisiveness is a constant characteristic unless they have structured guidelines. They fear making mistakes or misjudgements, and delay repeatedly until they have enough data to take what they consider the only right decision. When rules do not dictate the correct answer to a problem or when procedures for tasks are not laid down, decision-making or task initiation may become a lengthy and painful process.

People with this personality disorder are characterized by excessive conscientiousness and scruples. They are inflexible about matters of morality, ethics, or values. Moral principles and standards of performance have to be followed rigidly, and respect for authority and rules is absolute. Failure to do these things leads to irritation, anger, and self-criticism.

These people are stingy and mean, and often live with standards far below their actual socio-economic status. They dislike spending, believing that money should be saved in case of future difficulties. They have great difficulty in discarding worn-out or worthless objects, believing that they might be useful some day. They may hoard objects such as newspapers or broken appliances, even when they have no sentimental value.

These people are humourless and lack spontaneity of emotional expression. Usually they do not express anger directly. However, they are often angry in situations in which they are unable to control the behaviour of themselves or others. Anger is generally manifested by indirect aggressive acts (such as leaving a small tip or not providing minor help when expected). Their management of anger is closely related to their attitude of dominance–submission towards authority figures. They may be excessively submissive to a person in authority whom they respect, but obstructive with an authority figure whom they do not respect.

The affect of the obsessive person is controlled and stilted. It is not flat or blunted, but constricted. They do not laugh or cry, and feel uncomfortable with people who express their feelings. Their mood is usually serious but may appear anxious or depressed. In a clinical interview they may sit in a stiff unnatural posture, and seldom make spontaneous comments about their emotions. They usually relate their history in a pedantic and circumstantial manner. If interrupted by a question from the doctor, they have to finish their monologue before answering. When asked about feelings, they answer with lists of facts and circumstances. They can label emotions and feelings, but are unable to display them.

In summary, obsessive personalities love order, neatness, and sameness, and hates novelty, spontaneity, and change. They need control, security, and certainty, and avoid creativity, art, and excitement. They mitigate anxiety by following strict rules and repress emotional expression by avoiding spontaneity. They fear their inner fragile and aggressive emotional world.

(d) Course

Like other personality disorders, obsessive–compulsive personality disorder is present in early adulthood and tends to be persistent and constant. However, some adolescents with marked obsessive traits become warm, loving, and tender adults. On the other hand, intense obsessional traits in adolescence are occasionally a premorbid stage of schizophrenia ('pseudoneurotic schizophrenia'). The developmental relationship between obsessive–compulsive personality disorder and obsessive–compulsive disorder is controversial. In the past, it was suggested that most obsessive–compulsive personality disorder evolved to a full obsessive–compulsive disorder, indicating that the two syndromes were expressions of the same basic disorder. More recent investigations[55] indicate that most obsessive–compulsive disorder patients do not have a comorbid obsessive–compulsive personality disorder. A variety of psychiatric disorders may present in a patient with obsessive personality, but depressive and anxiety disorders are the most common, followed by phobic, somatoform, and obsessive–compulsive symptoms. Hypochondriacal syndromes are commonly found in obsessive individuals when they lose control of situations.

Persons with this personality disorder may do well in jobs that demand working with detail, order, and structured procedures, and may adjust to interpersonal relationships with submissive spouses. However, they are particularly vulnerable to unexpected changes in their occupational and social environment. Late-onset depression is a common occurrence in obsessive–compulsive personalities.

(e) Differential diagnosis

The main difficulty in diagnosing obsessive–compulsive personality disorder is to differentiate it from obsessive–compulsive disorder. The latter diagnosis is made when occupational and personal functioning is severely impaired as a consequence of doubt, indecisiveness, hoarding, or any other obsessive behaviour. In many, but not all, cases of obsessive personality, the traits and behaviours

are egosyntonic and no resistance is present, in contrast with obsessive–compulsive disorder.

The perfectionism of obsessive personalities may be present in narcissistic personality disorder. However, narcissistic individuals tend to believe that they have achieved perfection, while obsessive individuals tend to be highly critical of their own achievements.

Social detachment and the lack of empathy and warmth may suggest schizoid personality disorder. However, obsessive individuals constrain their emotional expression to keep control of a situation, while schizoids lack the fundamental capacity for affective display or intimacy.

Not all individuals with obsessive traits have obsessive–compulsive personality disorder. Obsessive traits can be adaptive in some situations; it is only when they are maladaptive, inflexible, and persistently cause functional impairment that a personality disorder be diagnosed.

(f) Treatment

Pharmacological treatment may be tried in patients with anxiety and distress due to intense doubts, indecisiveness, and scruples. Benzodiazepines may alleviate tension in these cases. Antidepressants with a serotonergic profile sometimes improve mood and global functioning.

Psychological cognitive treatment, focusing on perfectionism, rigidity, scrupulousness, and intolerance of failure, is the main therapeutic approach. Repressed aggression, guilt, and dependency needs should be addressed using a psychodynamic approach.

Other personality disorders (not included in DSM-IV)

Passive–aggressive (negativistic) personality disorder

(a) Definition

Resistance to demands for adequate social and occupational performance and negativistic attitude are considered to be central features of passive–aggressive personality disorder. A pervasive pattern of argumentativeness, oppositional behaviour, and defeatist attitudes are typical, and are thought to be a covert manifestation of underlying aggression, which is expressed passively and indirectly. Passive–aggressive personalities have interpersonal and cognitive dysfunction and severe impairment in terms of self-image, global mental health, and ability to function at work and in intimate relationships.[56]

Passive–aggressive personality disorder was officially included in DSM-I as the passive–aggressive personality 'trait disturbance' depicted as an immature reaction to military stress by helpless, passive, and obstructive resistant behaviour. However, passive–aggressive disorder was not included in DSM-IV because of the many unsolved problems related to its concept in previous classifications. Instead, it is placed in Appendix B of DSM-IV where it is alternatively called negativistic personality disorder. Research criteria are proposed which are expected to be empirically evaluated and to determine the validity and reliability of this diagnosis (Table 4.12.3.15).

There has been much debate as to whether passive aggression constitutes a personality disorder, a defence mechanism, or a specific maladaptive personality trait (coping style).[57] Surprisingly, empirical literature on the subject is scarce, although passive–aggressive behaviour has been widely recognized by clinicians. An overlap with other personalities has been shown, and it has never been included as a separate category in the *International Classification of Diseases*. The passive–aggressive dimension, as assessed by self-reports, is always high in depressed patients and is state-dependent.[58] Perhaps it would be best to conceptualize passive aggression as a continuum: a passive–aggressive defence mechanism may be normal in some situations, it could be a trait of many personality disorders, and when pronounced and long-lasting it should be designated as passive–aggressive personality disorder.

Table 4.12.3.15 DSM-IV research criteria for passive–aggressive personality disorder

A.	A pervasive pattern of negativistic attitude and passive resistance to demands for adequate performance, beginning by early adulthood and present in variety of contexts, as indicated by four (or more) of the following
	1 Passively resists fulfilling routine social and occupational tasks
	2 Complains of being misunderstood and unappreciated by others
	3 Is sullen and argumentative
	4 Unreasonably criticizes and scorns authority
	5 Expresses envy and resentment toward those apparently more fortunate
	6 Voices exaggerated and persistent complaints of personal misfortune
	7 Alternates between hostile defiance and contrition
B.	Does not occur exclusively during major depressive episodes and is not better accounted for by dysthymic disorder

(b) Epidemiology

The population prevalence ranges from 0.9 to 3 per cent, but in those cases in which a secondary co-occurring diagnosis was assigned, the secondary frequency of passive–aggressive personality disorder was about 10 per cent.[59] Some studies found a higher prevalence in women and others in men.

(c) Aetiology

The cause of the disorder is multidimensional, comprising biological, psychoanalytical, behavioural, interpersonal, and social learning perspectives.

Ambivalence is considered to be a core conflict of passive–aggressive personalities, which originates from fixation to the biting or sucking stages of the oral phase of psychosexual development. Some authors consider masochism to be another precursor of the passive–aggressive personality.

According to the behavioural model, passive–aggressive behaviour is the expression of anger in maladaptive verbal and non-verbal ways that do not lead to rewarding problem-solving Failure to learn appropriate assertive behaviour would be the main aetiological factor.[60]

The primary social factor influencing the development of passive–aggressive patterns would be contradictory parental attitudes in childhood which, being conflicting and incompatible, prevent the child from expressing his feelings directly and thus urge him to develop passive resistance.

(d) Clinical characteristics and diagnosis

Passive–aggressive personalities seek novel and stimulating situations in impulsive ways, while remaining unpredictable.[61] Procrastination and inefficiency are behaviours used to avoid responsibility, which they show by stubborn resistance to the fulfilment of expectations and claiming forgetfulness.

Passive–aggressive individuals easily become irritable and, gloomy and they are resentful and discontent with life. Accumulated anger may be expressed by verbal acting-out, after which passive–aggressive individuals feel guilty and gloomy. Since they have difficulties in expressing emotions directly, they are prone to diffuse somatic complaints, hypochondriasis, and psychosomatic disorders.

Interpersonal ambivalence is a core feature of passive–aggressive personality disorder. Negativism is particularly expressed towards authorities, with whom they are never satisfied and whom they criticize constantly. The argumentative self-detrimental behaviour of passive–aggressive personalities is often experienced by others as punitive and manipulative. Negative verbal comments, which are often caustic, and irritable and moody patterns of communication are typical.

Passive–aggressive individuals are cynical, sceptical, hypercritical, and mistrustful. Disillusioned with life, discouraged, discontented with themselves and with others, they are also pessimistic about the future. They persistently complain and blame others for their own bad luck, feeling themselves to be misunderstood martyrs and victims of destiny.

(e) Comorbidity and differential diagnosis

Passive–aggressive personality disorder is frequently comorbid with major depression, dysthymic disorder, anxiety, panic disorder, and hypochondriasis. Patients with depressive disorders are more aware of their feelings of inferiority and more likely to feel guilty, and their depressed mood is continuous rather than erratically hostile and moody, as in passive–aggressive personality.

Comorbidity with many personality disorders (histrionic, borderline, obsessive–compulsive, dependent, narcissistic) is also frequently observed. People with these personality disorders may use passive aggression as a defence mechanism. Suicide attempts are not as frequent as in histrionic and borderline personality disorders, and features of passive–aggressive personality are less dramatic, affective, openly aggressive, and severe.

(e) Course and prognosis

There are insufficient data on the course and prognosis of passive–aggressive personality disorder. When passive–aggressive people are unable to control their anger, they may experience anxiety, panic states, depressive episodes, chronic depression, and psychosomatic disorders. They are prone to alcohol abuse, and their careers are erratic and stunted despite their abilities (frequent changes of jobs are common). Suicide attempts may complicate this disorder.

(f) Treatment

(i) Pharmacotherapy

Target symptoms, such as depression, anxiety, and somatic complaints, should be treated. Benzodiazepines may be warranted and helpful in the initial period of psychotherapy. The side effects of medication are often the reason for complaints about their psychiatrist's inefficiency. Non-compliance is frequent, reflecting resistance to the therapist. Abuse of medicaments should be controlled and considered seriously.

(ii) Psychotherapy

Psychotherapy is the treatment of choice. Various modes are used, including supportive, psychodynamic, behavioural assertiveness training, and the paradoxical approach in group or individual settings. The goal of treatment should be to help the patient escape from the vicious circle of self-defeating behaviour and develop a mature way of expressing anger and other feelings.

Self-defeating (masochistic) personality disorder

Individuals with masochistic personality disorder persistently seek humiliation and failure, and submit to the will of others.

The term masochism was introduced to psychiatry by Krafft-Ebing in 1882. It is derived from a character in a novel by the German author Leopold von Sacher-Masoch, a man who endured torture, scorn, and humiliation from a woman. Later, Freud conceptualized masochism as the result of aggressive instinct directed towards the self instead of an external object. Reich[62] described the masochistic character as a person who suffered deep frustrations in early developmental stages and needed to express this frustration through suffering inflicted by the love 'objects'. Thus defiance is always present in the masochistic search for love. According to Horney,[63] helplessness and victimization may be a masochistic way of expressing hostility by making others feel shameful. Masochistic suffering may also be used to avoid reproach and responsibility and as a way of restoring a sign of personal value.

(a) Clinical picture

People with self-defeating personality disorder avoid pleasurable experiences and undermine their achievements. They neglect their appearance and live below their means. They accept and endure humiliation, expecting that others will sympathize with them. In this way, they fulfil their expectation that submission will bring love and care. They prefer to relate to people who abuse them and consider those who consistently treat them well to be boring.

These people fail to accomplish tasks of which they are capable and adopt an inferior role. When making appropriate demands, they feel that they are taking advantage of others and adopt an apologetic manner. For them, success is inversely related to inner security. Successful relationships do not make them feel confident, but increase their fears. They tend to believe that all experiences involve future frustration and pain. They may respond to positive personal events with depression and behaviour that negates their accomplishments. They look for people who will respond to their behaviour with disdain, rejection, or cruelty.

People with self-defeating personality disorder do not defend themselves against expressions of disgust and resentment directed towards them and rarely accuse or reproach others. They do not feel confident and are not assertive. They fear that optimism may lead to greater problems.

They are not worried by these attitudes. Rather, masochists believe that by exaggerating their weaknesses and inefficiency they will protect themselves from aggression by others. They feel protected when someone needs something from them, and many non-assertive masochists engage in self-sacrifice for their own protective feelings rather than for the welfare of others. What seems to be a comprehensive and self-sacrificing attitude reflects a lack of confidence and empathy.

The mood of these individuals is usually dysphoric, fluctuating between anxiety and sadness.

(b) Differential diagnosis

Since masochistic personality disorder is not included in DSM-IV, little research has been done on comorbidity or on the validity or specificity of this diagnosis.

Self-defeating attitudes, low self-esteem, and depression may be found in individuals with dependent, borderline, and depressive personality disorders. Some of the characteristics are also found in avoidant personality disorder.

Masochists are prone to mood disorder and dysthymia. Anxiety disorders are frequent. Their fears of abandonment are a persistent source of anxiety. Hypochondriasis and somatization are also common, sometimes as a way of obtaining attention.

(c) Treatment

Antidepressants and anxiolytics may be useful to alleviate a dysphoric state. Psychological treatment should take into account that masochistic patients sometimes induce an aggressive countertransference as a response to their own wish to be hurt. The therapist should gradually clarify the behaviours, which provoke hostile responses and seek to reward adaptive interpersonal behaviours. Training in assertiveness and social skills is sometimes helpful.

Sadistic personality disorder

Sadistic personality disorder is a controversial category, which is not included in either DSM-IV or ICD-10. Some authors, especially those working with perpetrators of abuse, support its inclusion in the diagnostic nomenclature, believing that it is a valid clinical entity, which deserves special treatment. Sadism as a term describing desire to inflict pain upon the sexual object was originally used by Krafft-Ebing.[64] Kernberg[32] connects two dispositions (sadistic and masochistic) into a sadomasochistic character, which includes 'help-rejecting complainers' and often has underlying borderline personality organization.

(a) Aetiology

Subjects with sadistic personality disorder often had a history of significant childhood loss and physical, emotional, or sexual abuse during childhood.[65] Despite their significant psychopathology, sadistic individuals are surprisingly highly functioning, with steady employment and intense long-lasting relationships.

(b) Clinical picture

Sadistic individuals demonstrate a long-standing maladaptive pattern of cruel, demeaning, and aggressive behaviour towards others in order to cause suffering and pain and to establish dominance and control. Sadistic persons are fascinated by violence, weapons, injuries, and torture, and are most frequently encountered in forensic settings among child and partner abusers.

(c) Differential diagnosis

The major distinction for the diagnosis of sadistic personality disorder is from antisocial personality disorder. Sadistic persons may simply represent an aggressive (antagonistic) subtype of psychopathy. Intimidation and sadistic control of others, as well as a fascination with weapons, martial arts, and torture, may be manifested by both antisocial and sadistic individuals. Moreover, both disorders may display 'malignant narcissism',[37] with an admixture of narcissistic, antisocial, sadistic, and paranoid features.

(d) Epidemiology

Existing data[66] suggest that sadistic personality disorder is relatively uncommon, although it may have a higher prevalence in specific forensic populations. Several studies found a high overlap with narcissistic, paranoid, and antisocial personality disorders, which raised the possibility that sadistic personality disorder may not be a distinct entity.

(e) Course

No data are available on the course of this disorder.

(f) Treatment

Sadistic individuals seldom seek treatment, and are usually encountered in forensic settings. No treatment has been successful for this disorder. Since the aetiology is probably multifactorial, there should be multiple approaches to treatment. The primary aim of treatment is control of cruelty and malignant aggression. As with antisocial personalities, selective serotonin reuptake inhibitors and lithium may be beneficial in regulating the serotonergic function that probably underlies the aggressive action, and carbamazepine and clonazepam may act to regulate ictal aggressive outbursts.

Psychotherapy is usually difficult because these individuals lack any desire to change and because there may be serious countertransference problems for the therapists. Small groups may be helpful because there is a dissolution of transference and peer confrontation is accepted more easily.

Depressive personality disorder

This personality disorder is not included in ICD-10 or DSM-IV, although it is considered as a subject for further study in the latter. However, the concept of depressive personality was well recognized in previous decades (e.g. by Kraepelin and Schneider). Depressive personality was seen as a pattern of brooding, pessimism, and low self-confidence,[2] and as a tendency to physical lassitude and suffering. Phenomenological accounts[67] emphasize dependency, orderliness, and adherence to social conventions. The psychoanalytic concept of masochism[47] overlaps to some extent with the classical depressive personality. More recently, Akiskal[68] has discussed the depressive personality as part of a spectrum of affective disorders, reviving Kretschmer's ideas[10] on the role of temperament as the base from where psychiatric disorders develop.

(a) Aetiology

Psychological and biological factors have been suggested as causes of depressive personality disorder. Early losses, inadequate attention from parents, and a punitive superego have been postulated by psychoanalytic authors. Others have suggested that a depressive temperament is genetically related to affective disorder.

(b) Clinical picture

People with depressive personality disorder are submissive, quiet, introverted, and unassertive. Their cognitive style is marked by pessimism, dejection, and self-reproach. They appear gloomy, joyless, cheerless, and unhappy. They are serious and lack a sense of humour. They do not believe that they deserve to be happy. They have a negative view of the past and present and do not expect things to improve. They anticipate failure and dwell on their negative perspective of life. They have a low tolerance to shortcomings and failures, which are seen as confirming their own pessimistic assumptions. They are prone to guilt and judge themselves severely.

Their self-esteem is low and they feel inadequate. They focus on the failings of others and are critical of themselves.

(c) Epidemiology

No data are available on the frequency of this disorder in the population.

(d) Differential diagnosis

Some clinicians doubt whether a distinction can be made between depressive personality disorder and dysthymic disorder. The diagnosis of depressive personality disorder emphasizes cognitive and behavioural aspects rather than the depressed mood. Also, dysthymia, although chronic, has a fluctuating course.

Dysthymic patents generally experience their symptoms as egodystonic, while people with depressive personality disorder think that they have a realistic view of their situation. People with depressive personality disorder may meet the criteria for dependent personality disorder and self-defeating personality disorder, and it is difficult to distinguish between these disorders.

(e) Course

Patients with depressive personality disorder have depressive episodes and dysthymia more frequently than other individuals, and may have difficulties in adapting to stressful or uncertain situations.

(f) Treatment

Antidepressants may be useful when the person is at higher risk for developing a depressive episode. Psychological treatment may be helpful, since these people have a good capacity for introspection and reality testing. Cognitive approaches can help patients to understand their negative views and derogatory attitudes. These patients usually establish a good therapeutic relationship with the clinician.

Personality changes

Enduring personality changes after traumatic experiences

ICD-10 has two categories for personality changes: those occurring after catastrophic situations, and those starting after psychiatric illness. Either diagnosis should be made only when there is evidence of a definite personality change, including cognition, behaviour, and interpersonal relationships. The changes must not be a manifestation of a current mental disorder or the residue of a previous mental disorder. It must not be an exacerbation of a pre-existing personality disorder.

The **aetiology** of the personality change presumably relates to the extreme existential experience of the catastrophic situation or the psychiatric illness.

Examples of catastrophic experiences include life in concentration camps, experiences of disaster, and prolonged exposure to other life-threatening situations. Personality changes following short-term exposure to life-threatening situations, such as a car accident, should not be included in this group, since such changes probably depend on a previous psychological vulnerability. In ICD-10, the symptoms of personality change after catastrophic experiences include hostility and mistrust of the environment, social withdrawal, feelings of emptiness or hopelessness, estrangement, and alertness or feeling on the edge.

When the personality change follows a mental disorder, the cause is related to the stressful experience and the perceived damage to the patient's self-image arising from the disease. Other people's attitudes towards the illness, the subjective emotional experience, and previous psychological adjustment are also important. Features of the personality change after mental disorders include feelings of being stigmatized and consequent withdrawal, passivity, and loss of interests, dependence, excessive demandingness and complaining, and dysphoric–labile mood.

Personality change due to a general medical condition (JLC)

This category, which is included in both ICD-10 and DSM-IV, describes syndromes affecting global features of behaviour, cognition, and emotions, and secondary to the physiological effects of general medical diseases.

The essential feature is a change in personality after a general medical disease. In childhood before a stable pattern of personality has been established, a marked deviation from normal development suggests the diagnosis.

A central feature is loss of control over the expression of emotions and impulses. Affect is commonly labile and shallow, although persistent mild euphoria or apathy may be present, especially when the frontal lobes are affected. The elevated mood of these patients is hollow and silly, unlike that of hypomania. Patients may appear childish, expansive, and disinhibited, but they may admit to not feeling happy. Others are indifferent and apathetic.

Exaggerated expressions of rage and aggression are usually present, often out of proportion to any stressor. Loss of impulse control is also shown in social and sexual disinhibition, inappropriate jokes, and overeating.

(a) Aetiology

In most cases this disorder is associated with structural damage to the brain. Head trauma, cerebral neoplasms, vascular accidents, multiple sclerosis, Huntington's disease, and complex partial epilepsy may all cause personality change, especially when affecting frontal and temporal lobes. Systemic diseases involving the central nervous system, including endocrine disorders, AIDS, lupus erythematosus and chronic metal poisoning, may have the same effect.

Patients with this disorder generally have a clear sensorium but may be inattentive and have some mild cognitive dysfunction. They do not show intellectual deterioration.

(b) Differential diagnosis

In dementia, personality change is accompanied by intellectual deterioration. However, personality change may predate the dementia. Distinction from schizophrenia and other personality disorders is based on the clinical history and the presence of a general medical disease capable of causing personality change.

(c) Treatment

If possible, treatment is directed to the causative condition. Pharmacological treatment of specific symptoms may be useful when depression or inappropriate anger is present.

Further information

Millon, T., and Davis, R. (2000). *Personality disorders in modern life*. John Wiley and sons, Inc., New York.

Schneider, K. (1950). *Psychopathic personalities* (9th edn.). Casell, London. (Original work published in 1923).

Menninger, K. (1940). Character disorders. In *The psychodynamics of abnormal behaviour* (ed. J.F. Brown), pp. 384–403. McGraw-Hill, New York.

Livesley, W.J., Jackson, D.N., and Schroeder, M.L. (1992). Factorial structure of traits delineating personality disorders in clinical and general population samples. *Journal of Abnormal Psychology*, **101**, 432–40.

References

1. Kraepelin, E. (1921). *Manic-depressive insanity and paranoia*. Lingstone, Edinburgh.
2. Schneider, K. (1950). *Psychopathic personalities* (9th edn.). Cassell, London.
3. Freud, S. (1911). Psychoanalytic notes upon an autobiographical account of a case of paranoia. In *Collected papers*, Vol. 3. Hogarth Press, London, 1925.
4. Perry, J.C., and Vaillant, G.E. (1989). Personality disorders. In *Textbook of psychiatry* (eds. H.I. Kaplan and B.J. Sadock), pp. 1352–87. Williams and Wilkins, Baltimore, MD.
5. Kendler, K.S., Masterson, C.C., Ungaro, R., *et al.* (1984). A family history study of schizophrenia related personality disorders. *The American Journal of Psychiatry*, **141**, 424–7.
6. Seiver, L.J. and Davis, K.L. (1991). A psychobiological perspective on the personality disorders. *The American Journal of Psychiatry*, **148**, 1647–58.
7. Millon, T. (1997). *Disorders of personality: DSM-IV and beyond* (2nd edn.). Wiley, New York.
8. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn.). American Psychiatric Association, Washington, DC.
9. Sullivan, H.S. (1953). *The interpersonal theory of psychiatry*. Norton, New York.
10. Kretschmer, E. (1936). *Physique and character*. Kegan Paul, Trench, Trubner, London.
11. American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd edn. revised). American Psychiatric Association, Washington, DC.
12. Robins, L.N. (1966). *Deviant children grown up: a sociological and psychiatric study of sociopathic personality*. Williams and Wilkins, Baltimore, MD.
13. de Girolamo, G. and Reich, J.H. (1993). *Personality disorders*. WHO, Geneva.
14. Crowe, R.R. (1974). An adoption study of antisocial personality. *Archives of General Psychiatry*, **31**, 785–91.
15. Cadoret, R.J. (1978). Psychopathology in adopted-away offspring of biologic parents with antisocial behaviour. *Archives of General Psychiatry*, **35**, 176–84.
16. Coccaro, E.F., Siever, L.J., Klar, H.M., *et al.* (1989). Serotonergic studies in patients with affective and personality disorders. *Archives of General Psychiatry*, **46**, 587–99.
17. Dee Higley, J., Mehlman, P.T., Taub, D.M., *et al.* (1992). Cerebrospinal fluid monoamine and adrenal correlates of aggression in free-ranging rhesus monkeys. *Archives of General Psychiatry*, **49**, 436–41.
18. Glueck, B. and Glueck, E. (1956). *Physique and delinquency*. Harper, New York.
19. Gunderson, J.G. (1984). *Borderline personality disorder*. American Psychiatric Press, Washington, DC.
20. Akhtar, S. (1992). *Broken structure*. Aronson, Northvale, NJ.
21. Jones, M. (1952). *Social psychiatry: a study of therapeutic communities*. Tavistock Publications, London.
22. Stern, A. (1938). Psychoanalytical investigation and therapy in borderline group of neuroses. *Psychoanalytic Quarterly*, **7**, 467–89.
23. Gunderson, J.G. and Kolb, J.E. (1978). Discriminating features of borderline patients. *The American Journal of Psychiatry*, **135**, 792–6.
24. Kernberg, O. (1989). The narcissistic personality disorder and the differential diagnosis of antisocial behaviour. *The Psychiatric Clinics of North America*, **12**, 553–70.
25. Quality Assurance Project. (1991). Treatment outlines for borderlines, narcissistic and histrionic personality disorders. *The Australian and New Zealand Journal of Psychiatry*, **25**, 392–403.
26. Akiskal, H.S. (1981). Subaffective disorders: dysthymic, cyclothymic and bipolar II disorders in the 'borderline realm'. *The Psychiatric Clinics of North America*, **4**, 25–46.
27. Widiger, T.A. and Frances, A. (1989). Epidemiology, diagnosis, and comorbidity of borderline personality disorder. In *American PsychiatricPress review of psychiatry*, Vol. 8 (eds. A. Tasman, R.E. Hales, and A.J. Frances), pp. 8–24. American Psychiatric Press, Washington, DC.
28. Stone, M.H. (1980). *The borderline syndromes. Constitution, personality and adaptation*. McGraw-Hill, New York.
29. Coccaro, E.F., Siever, L.J., Klar, H.M., *et al.* (1989). Sertonergic studies in patients with affective and personality disorders: correlates with suicidal and impulsive aggressive behaviour. *Archives of General Psychiatry*, **46**, 587–99.
30. New, A.S., Hazlett, E.A., Buchsbaum, M.S., *et al.* (2007). Amygdala-prefrontal disconnection in borderline personality disorder. *Neuropsychopharmacology*, **32** (7), 1629–40.
31. Zanarini, M.C., Williams, A.A., Lewis, R.E., *et al.* (1997). Reported pathological childhood experiences associated with the development of borderline personality disorder. *The American Journal of Psychiatry*, **154**, 1101–6.
32. Kernberg, O.F. (1975). *Borderline conditions and pathological narcissism*. Aronson, New York.
33. Gunderson, J.G. (1996). The borderline patient's intolerance of aloneness: insecure attachment and therapist availability. *The American Journal of Psychiatry*, **153**, 752–8.
34. Akhtar, S. (1984). The syndrome of identity diffusion. *The American Journal of Psychiatry*, 141, 1381–5.
35. Gunderson, J.G. and Phillips, K.A. (1995). Personality disorders. In *Comprehensive textbook of psychiatry* (5th edn.) (eds. H.I. Kaplan and B.J. Sadock), pp. 1438–41. Williams and Wilkins, Baltimore, MD.
36. Stein, G. (1992). Drug treatment of personality disorders. *The British Journal of Psychiatry*, **161**, 167–84.
37. Kernberg, O.F. (1984). *Severe personality disorder-psychotherapeutic strategies*. Yale University Press, New Haven, CT.
38. Paris, J. (1994). *Borderline personality disorder-a multidimensional approach*. American Psychiatric Press, Washington, DC.
39. Linehan, M.M., (1987). Dialectical behavior therapy: a cognitive behavioral approach to parasuicide. *Journal of Personality Disorders*, **1**, 328–33.
40. Kraepelin, E. (1913). Hysterical insanity. In *Lectures on clinical psychiatry* (trans. T. Johnstone), pp. 249–58. William Wood, New York.
41. Freud, S. (1931). Libidinal types. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 21 (ed. J. Strachey), p. 266. Hogarth Press, London, 1961.
42. Zimmerman, M. and Coryell, W.H. (1990). Diagnosing personality disorders in the community. *Archives of General Psychiatry*, **47**, 527–31.
43. Marmor, J. (1953). Orality in the hysterical personality. *The Journal of the American Psychoanalytic Association*, **1**, 656–71.

44. Torgersen, S. (1980). The oral, obsessive and hysterical personality syndromes. A study of hereditary and environmental factors by means of the twin method. *Archives of General Psychiatry*, **37**, 1272–7.
45. Cadoret, R.J. (1978). Psychopathology in adopted-away offspring of biologic parents with antisocial behaviour. *Archives of General Psychiatry*, **35**, 176–84.
46. Zetzel, E. (1968). The so-called good hysteric. *The International Journal of Psychoanalysis*, **49**, 256–60.
47. Kernberg, O. (1988). Hysterical and histrionic personality disorders. In *The personality disorders and neuroses* (eds. A. Cooper, A. Frances, and M. Sacks), pp. 231–41. J.B. Lippincott, Philadelphia, PA.
48. Zimmerman, M. and Coryell, W.H. (1990). Diagnosing personality disorders in the community. *Archives of General Psychiatry*, **47**, 527–31.
49. Ronningstam, E. and Gunderson, J. (1991). Differentiating borderline personality disorder from narcissistic personality disorder. *Journal of Personality Disorders*, **5**, 225–32.
50. Millon, T. (1981). *Disorders of personality: DSM-III Axis II*. Wiley, New York.
51. Cloninger, C.R. (1986). A unified biosocial theory and its role in the development of anxiety states. *Psychiatric Development*, **3**, 167–226.
52. World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders-clinical descriptions and diagnostic guidelines*. World Health Organization, Geneva.
53. Clifford, C.A., Murray, R.M., and Fulker, D.W. (1984). Genetic and environmental influences on obsessional traits and symptoms. *Psychological Medicine*, **14**, 791–800.
54. Erikson, E. (1959). Growth and crises of the healthy personality. In *Psychological issues* (ed. G.S. Klein). International Universities Press, New York.
55. Baer, L. and Jenike, M.A. (1992). Personality disorders in obsessive compulsive disorder. *The Psychiatric Clinics of North America*, **15**, 803–12.
56. Drake, R.E. and Vaillant, G.E. (1985). A validity study of axis II of DSM-III. *The American Journal of Psychiatry*, **142**, 553–8.
57. Perry, J.C. and Flannery, R.B. (1982). Passive-aggressive personality disorder: treatment implications of a clinical typology. *The Journal of Nervous and Mental Disease*, **170**, 164–73.
58. Lecic-Tosevski, D. and Divac-Jovanovic, M. (1996). Effects of dysthymia on personality assessment. *European Personality*, **11**, 244–8.
59. Millon, T. (1993). Negativistic (passive-aggressive) personality disorder. *Journal of Personality Disorders*, **7**, 78–85.
60. Millon, T. (1981). *Disorders of personality. DSM-III: Axis II*. Wiley, New York.
61. Cloninger, C.R. (1987). A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry*, **44**, 573–88.
62. Reich, W. (1933). *Character analysis* (3rd edn.) (trans. V.R. Carfagno). Farrar, Straus and Giroux, New York.
63. Horney, K. (1945). *Our inner confl icts*. Norton, New York.
64. Krafft-Ebing, R. (1989). *Psychopathia sexualis* (10th edn). Enke, Stuttgart.
65. Gay, M. and Fiester, S. (1991). Sadistic personality disorder. In *Psychiatry* (ed. R. Michaels). J.B. Lippincott, Philadelphia, PA.
66. Fiester, S.J. and Gay, M. (1995). Sadistic personality disorder. In *The DSM-IV personality disorders* (ed. W.J. Livesley), pp. 329–40. Guilford Press, New York.
67. Tellenbach, H. (1980). *Melancholia*. Duquesne University Press, Pittsburgh, PA.
68. Akiskal, H.S. (1989). Validating affective personality types. In *The validity of psychiatric diagnosis* (ed. L. Robins), pp. 217–27. Raven Press, New York.

4.12.4 Epidemiology of personality disorders

Francesca Guzzetta and Giovanni de Girolamo

Introduction

tbegun to be scientifically investigated. This development has taken place because a number of standardized instruments to assess personality and PD in an empirical fashion have been developed, in parallel with the refinement of a valid and reliable diagnostic system based on a categorical approach.

The need for the epidemiological investigation of PDs seems justified for several reasons.

1 As seen in recent epidemiological surveys, PDs are frequent and have been found in different countries and sociocultural settings.

2 PDs can seriously impair the life of the affected individual and can be highly disruptive to societies, communities, and families.

3 Personality status is often a major predictive variable in determining the outcome of Axis I mental disorders and the response to treatment.

In this chapter, we review the epidemiological literature on PDs up to October 2007, focusing on studies carried out since the development of the DSM-III. First, community prevalence studies of PDs are reviewed. We then look at the prevalence of individual PDs in the community. Finally, we consider the prevalence of PDs in clinical populations, and in special settings (e.g. primary care, prisons, etc.).

Community epidemiological studies of unspecified personality disorders

Until the development of the DSM-III diagnostic criteria for PDs and the subsequent availability of standardized assessment instruments, epidemiological studies aimed at assessing the prevalence rate of PDs were hampered by severe methodological limitations, including differences in sampling methods and in diagnostic criteria, the known unreliability of PD diagnoses based on clinical judgement, and the lack of standardized assessment methods. Since 1980, twelve main studies with at least 200 subjects sampled have ascertained the prevalence rate of PDs in different community samples using assessment instruments specific for PD; they are shown in Table 4.12.4.1.

In these studies, the sample sizes ranged between 200 and 2053 subjects, with an average sample of 565.4; all surveyed individuals were evaluated by means of a specific PD assessment instrument, mainly a structured interview. While most studies were carried out in one stage, Lenzenweger *et al.*[6] first screened a large sample of university students with a self-administered Axis II inventory, and then interviewed a subgroup of 258 subjects using the International Personality Disorder Examination. The median prevalence rate of any PDs in these eight studies is 12.5 per cent.

Two large community studies[13,14] carried out in the USA were not included in Table 4.12.4.1 since PD prevalence rates were based

Table 4.12.4.1 Prevalence rates of personality disorders in epidemiological surveys

Reference	Country	Sample size	Sample features	Diagnostic Criteria	Assessment method	PD prevalence rate (%)
Black *et al.*[1]	USA	247	Relatives of obsessive-compulsive and normal control probands	DSM-III-R	SIPD	**22.3**[a]
Casey & Tyrer[2]	UK	200	Urban and rural residents	ICD-9	PAS	**13.0**
Coid *et al.*[3]	UK	626	Urban and rural residents aged 16–74 and selected in a 2 phase survey (weighted data)	DSM-IV	SCID-II	**4.4**
Crawford *et al.*[4]	USA	644	Individuals re-interviewed from previous surveys, mean age 33.1 years (range 27.7–40.1)	DSM-IV	SCID-II	**15.7**
Klein *et al.*[5]	USA	229	Relatives of normal controls	DSM-III-R	PDE	**14.8**
Lenzenweger *et al.*[6]	USA	258	University students age 18–19 years (two-stage procedure)	DSM-III-R	IPDE	**3.9**[b]
Maier *et al.*[7]	Germany	452	Normal controls, their partners, and relatives	DSM-III-R	SCID-II	**10.0**
Moldin *et al.*[8]	USA	302	Normal controls, parents and their children	DSM-III-R	PDE	**7.3**
Reich *et al.*[9]	USA	235	Urban residents	DSM-III	PDQ	**11.1**
Samuels *et al.*[10]	USA	742	Individuals re-interviewed from previous survey, aged 34–94 years (weighted data)	DSM-IV	IPDE	**9.0**
Torgersen *et al.*[11]	Norway	2,053	Individual from National Register (weighted data)	DSM-III-R	SIPD	**13.4**
Zimmerman & Coryell[12]	USA	797	Relatives of patients and normal controls	DSM-III	SIDP	**14.3**[a]

PAS, Personality Assessment Schedule, IPDE, International Personality Disorder Examination; PDE, Personality Disorder Examination; PDQ, Personality Diagnostic Questionnaire; SCID—II, Structured Clinical Interview for DSM-IV Axis II disorders; SIDP, Structured Interview for DSM-III-R Personality.

[a] Prevalence includes those with 'mixed' and 'not otherwise specified' disorder.

[b] Prevalence was 6.7% 'definite', 11% 'possible', including 'not otherwise specified disorder'.

on screening questions[13] and on a newly developed fully diagnostic structured interview carried out by lay interviewers rather than clinicians, which lacked any accompanying validity data.[14]

In the surveys considered here, the rate of PDs decreases in older age groups; although the sex ratio is different for specific types of PD (e.g. more schizoid, narcissistic, and antisocial PDs among males, more dependent, avoidant, and histrionic PDs among females), the overall rates of PD are about equal for both sexes. Finally, prevalence rates are generally higher in urban populations and lower socio-economic groups.

Community epidemiological studies of specified personality disorders

Table 4.12.4.2 lists the median prevalence rates for specified PDs based on studies that surveyed different types of randomly selected community samples. We will comment on some of the findings. The first column shows the number of studies on which the median prevalence rate is based.

Paranoid personality disorder

The median prevalence rate of paranoid PD is 1.6 per cent. In the study by Baron,[15] paranoid PD was remarkably more common among relatives of schizophrenic probands (7.3 per cent) than among relatives of control probands (2.7 per cent).

Schizoid personality disorder

There have been 13 studies evaluating the prevalence of schizoid PD in the community, with a median prevalence rate of 0.8 per cent. Baron[15] reported a rate of 1.6 per cent of schizoid PD among relatives of schizophrenic probands, but no cases among relatives of control probands.

Table 4.12.4.2 Median prevalence rates of specified personality disorders in epidemiological surveys

PD Category	Number of studies (*N*)	Median prevalence rate (%)
Paranoid	13	1.6
Schizoid	13	0.8
Schizotypal	13	0.7
Antisocial (dissocial)	24	1.5
Borderline	15	1.6
Histrionic	12	1.8
Narcissistic	10	0.2
Obsessive-compulsive	13	2.0
Avoidant (anxious)	13	1.3
Dependent	12	0.9
Passive-aggressive	8	1.7

Schizotypal personality disorder

The median prevalence rate of schizotypal PD is 0.7 per cent. However, in the study by Baron[15] schizotypal PD was remarkably more common among relatives of schizophrenic probands (14.6 per cent) than among relatives of control probands (2.1 per cent). In a similar fashion, Asarnow *et al.*[16] reported significantly higher rates of schizotypal personality disorders in relatives of probands with childhood onset of schizophrenia than in relatives of community controls (4.2 per cent vs. 0 per cent). These results provide additional support for the specific relationship between schizophrenia and schizotypal PD.

Antisocial (dissocial) personality disorder

Antisocial is the most studied PD. Its prevalence has been assessed in 24 epidemiological surveys, with a mean sample size of 2 943 subjects; nine studies used the Diagnostic Interview Schedule as assessment instrument. Antisocial PD seems to have a prevalence of around 1.5 per cent in the general population and to be substantially more frequent among males than females, with sex ratios ranging from 2:1 to 7:1. It is also more common among younger adults, those living in urban areas, and the lower socio-economic classes. People with a diagnosis of antisocial PD are also high users of medical services.

Borderline personality disorder

Borderline PD has been investigated in 15 studies. Swartz *et al.* [17] carried out a study among 1 541 community subjects (between 19 and 55 years of age) at the North Carolina site of the Epidemiologic Catchment Area (ECA) Program, using a diagnostic algorithm derived from the Diagnostic Interview Schedule (**DIS**). They found a rate of 1.8 per cent for borderline PD; the disorder was significantly more common among females, the widowed, and the unmarried. There was a trend towards an increase in the diagnosis in younger, non-white, urban, and poorer respondents. The highest rates were found in the 19 to 34 age range, with the rates declining with age. All borderline respondents had also a DIS DSM-III Axis I lifetime diagnosis.

Although some believe there is a preponderance of females with borderline PD, they do not always take into account that there is also a preponderance of females in the populations studied. There were two studies that did not find a higher female prevalence.[11,18]

Histrionic personality disorder

Histrionic PD has a median prevalence rate of 1.8 per cent, ranging from 0 per cent[3] to 3.2 per cent.[1] A study by Nestadt *et al.*[19] carried out at the Baltimore (Maryland) site of the Epidemiologic Catchment Area Program, ascertained the prevalence of histrionic PD in the community. The authors found a prevalence of 2.1 per cent in the general population, with virtually identical rates in men and women. No significant differences were found in terms of race and education, but the prevalence was significantly higher among separated and divorced persons. It should be noted that the study derived the diagnoses from instruments not originally intended to diagnose PDs; it might be possible that, in some cases, this study has identified personality traits rather than 'true' PDs.

Narcissistic personality disorder

No cases of narcissistic PD were found in five studies. However, Reich *et al.*[9] and Lenzenweger *et al.*[6] found rates of 0.4 per cent and 1.2 per cent respectively, even higher rates were found by Klein *et al.*[5] and Crawford *et al.*[4] who reported prevalence rates of 3.9 and 2.2 per cent respectively.

Obsessive compulsive (anankastic) personality disorder

The median prevalence rate of obsessive compulsive PD, obtained from 13 studies, was found to be 2 per cent, the highest of all PDs. The rate of compulsive PD was especially high in a study in which the Personality Diagnostic Questionnaire was used (6.4 per cent).[6] However, lower rates were reported with structured interviews. A community study, carried out at the Epidemiologic Catchment Area Program Baltimore site, found a prevalence of 1.7 per cent.[20] Males had a rate about five times higher than females. The disorder was also more frequent among white, highly educated, married, and employed subjects, and it was associated with anxiety disorders. However, the study derived the diagnosis from an interview originally not intended to diagnose PDs; this may mean that adaptive obsessive compulsive traits, rather than a 'true' PD, were identified. In the study by Black *et al.* [1] rates of obsessive compulsive PD were higher among relatives of probands with obsessive compulsive disorder compared to relatives of comparison probands (10.8 per cent vs. 7.9 per cent, respectively), however this difference did not reach statistical significance.

Avoidant (anxious) personality disorder

A total of 13 studies have investigated the prevalence of avoidant PD in community samples, with a median prevalence rate of 1.3 per cent. In the study by Asarnow *et al.*[16] avoidant PD occurred more frequently in relatives of schizophrenia probands compared to comparison control probands, also when controlling for schizoid or paranoid PD, and the authors suggest that avoidant PD might be a separate schizophrenia spectrum disorder, and not merely a sub-clinical form of schizoid or paranoid PD.

Dependent personality disorder

In 12 studies in which the frequency of dependent PD was assessed, the median prevalence rate was 0.9 per cent.

Passive-aggressive personality disorder

The median prevalence rate of passive-aggressive PD, obtained from 8 studies, was found to be quite high (1.8 per cent); interestingly, this type of PD has not been included either in DSM-IV or in ICD-10.

Epidemiological studies of personality disorders carried out in psychiatric settings

Table 4.12.4.3 lists the median prevalence rates for any PDs found in 61 studies carried out in inpatient and outpatient psychiatric samples and published between 1981 and October 2007. Only prospective studies that surveyed clinical samples (either inpatients or outpatients) of more than 100 subjects have been considered for

Table 4.12.4.3 Median prevalence rates of PDs among psychiatric patients in prospective studies including more than 100 subjects

Diagnostic category	Number of studies (*N*)	Median sample (*N*)	Median prevalence rates (%)
Alcohol and substance abuse	15	250	57.0
Affective disorder	19	200	49.2
Anxiety disorders	7	200	40.4
Any Axis disorder	20	131	51.0

this analysis. The second column shows the number of studies on which the median prevalence rate is based.

In these studies, subjects have been directly evaluated for the purpose of obtaining PD rates, by means of a standardized assessment instrument specific for PDs. Several other studies, which have evaluated only the prevalence of specified PDs in clinical samples, are not shown here.

In general, the prevalence of PDs among psychiatric outpatients and inpatients is quite high, with a substantial number of studies ($n = 29$) showing a PD prevalence rate equal or higher than 50 per cent of the sample. However, it is difficult to draw more definite conclusions from these studies, because of substantial differences in sampling, diagnostic criteria, timing of the assessment, assessment methods, availability of mental health services, prevalence of Axis disorders, and sociocultural factors.

There are, however, some consistencies across studies that deserve consideration. The most prevalent PD seems to be borderline, both in inpatient and in outpatient settings. The next most common PDs is histrionic, whereas schizoid PD is infrequently diagnosed. Borderline and histrionic PDs are also characterized by the lowest social functioning. They are especially common in inpatient settings, as their symptomatology often results in the patient being admitted to hospital due to their suicidal behaviour, substance abuse, and cognitive-perceptual abnormalities. In outpatient settings, dependent, and passive-aggressive PD are also common.

Especially in inpatient settings, many people who meet the criteria for one PD also meet the criteria for other PDs [21–23]. The highest comorbidity rate appears to occur with borderline PD, with the frequent coexistence of borderline and histrionic PDs, antisocial, schizotypal, and dependent PDs.

With regard to comorbidity between PDs and Axis I disorders, the most common and best-studied patterns are between substance abuse and PDs, affective disorders and PDs, and anxiety disorders and PDs (particularly borderline, antisocial, avoidant, and dependent PDs). Other clinically significant associations have been found between PDs and eating disorders: obsessive-compulsive, avoidant, and dependent PDs are most commonly associated with anorexia nervosa whereas borderline, avoidant, dependent, and paranoid PDs are the most common among individuals with bulimia.[24] High rates of PD (especially borderline and antisocial PDs) have also been detected in patients with selected medical conditions, such as HIV-positive patients.[25,26]

Some studies have assessed the treated prevalence of PD using administrative data (e.g. discharge figures, psychiatric case register data, etc.). In the United States, using data from the 1993 National Hospital Discharge Survey, Olfson and Mechanic[27] found that almost 12 per cent of patients discharged from public general hospitals had a diagnosis of PD, compared with 11 per cent of patients from non-profit hospitals and 5 per cent of patients from proprietary general hospitals.

Some investigations, which compared the hospital admission rates for PD over time, allow us to assess the impact of diagnostic changes. In Denmark, sex- and age-standardized rates of first-admitted borderline patients significantly increased during the 16-year interval between 1970 and 1985, and this might be explained in terms of a change in diagnostic habits.[28] In the United States, comparing the diagnoses given to inpatients in a large university-affiliated mental hospital in the last 5 years of the DSM-II era ($n = 5143$) with those given in the first 5 years of the DSM-III era ($n = 5771$), a marked increase (from 19 per cent to 49 per cent) was found in the diagnosis of PD, together with a decrease in the diagnosis of schizophrenia and a corresponding increase in the diagnosis of affective disorders.[29]

The epidemiological findings in treated samples are especially important if we bear in mind that the presence of a PD among those suffering from other mental disorders can be a major predictor of the natural history and treatment outcome. Therefore, an important clinical implication of these findings is that patients in treatment because of severe Axis I disorders must be carefully assessed with an assessment instrument specific for PDs, because of the high likelihood of diagnosing a PD and the subsequent need to adjust their treatment accordingly.

Epidemiological studies of personality disorders carried out in other settings

A few epidemiological studies on PDs have been carried out among patients attending primary healthcare settings; in these studies between 5 and 8 per cent of patients have been identified as having a primary diagnosis of PD.[21] When the assessment is made independent of the primary diagnosis, however, the average prevalence rate can rise several-fold because of state effects. In a consecutive sample of primary care *attenders*, a PD was diagnosed in 24 per cent of the sample (N=303) and was associated with the presence of common mental disorders and unplanned surgery attendance, indicating that PDs may represent a significant source of burden in primary care.[30]

In other institutional settings, such as prisons, several studies have found very high rates of PDs. In the United Kingdom, two large-scale studies have been completed; in the first, carried out among 750 prisoners representing a 9 per cent cross-sectional sample of the entire male unconvicted population, a PD was diagnosed in 11 per cent of the sample.[31] In the second study, a representative sample of the entire prison population of England and Wales was evaluated; a sub-sample was assessed with the SCID-II administered by a clinician.[32] The prevalence rates for any PD were 78 per cent for male remand prisoners, 64 per cent for male sentenced prisoners, and 50 per cent for female prisoners. In a large meta-analysis by Fazel & Danesh [33] of 28 studies,

including a total of 13 844 prisoners, antisocial PD was diagnosed in 47 per cent of subjects. High rates of borderline and antisocial PDs have also been found in a sample ($n = 805$) of women felons entering prison in a North American State.[34]

Conclusions

Up to 30 years ago, the epidemiology of PDs had not received the same amount of attention as that of many other psychiatric disorders. Since then the situation has changed, and we now have data on the prevalence of PD in the community and in psychiatric facilities. Community data come primarily from 12 studies, with a total sample of 6 785 subjects from four countries (Germany, Norway, the United Kingdom, the United States). There are excellent national and cross-national epidemiological data on antisocial personality disorder based on the same diagnostic methods. There are almost no data on other PDs from countries other than the United States, the United Kingdom, Germany and Norway.

One important methodological problem is that some PDs have a very low prevalence rate. Consequently, epidemiological surveys carried out among the general population may require very large samples in order to identify a sufficient number of cases to study demographic correlates and the association of PD with other psychiatric disorders. Future studies should try to address this problem and provide us with more definite epidemiological data. These data will also be invaluable in showing the validity of current classifications and in better delineating the boundaries between different PDs.

Further information

The complete bibliography of studies included in tables 4.12.4.2–4.12.4.3 can be asked from the authors at the e-mail address: f.guzzetta@gmail.com.

Torgersen S. (2005). Epidemiology. In *The American Psychiatric Publishing Textbook of Personality Disorders* (eds. Oldham J.M., Skodol, A.E., Bender, D.S.), pp. 129–41. The American Psychiatric Publishing.

References

1. Black, D.W., Noyes, R., Jr., Pfohl, B., *et al.* (1993). Personality disorder in obsessive-compulsive volunteers, well comparison subjects, and their first-degree relatives. *American Journal of Psychiatry,* **150**(8) 1226–32.
2. Casey, P.R., Tyrer, P.J. (1986). Personality, functioning and symptomatology. *Journal of Psychiatric Research,* **20**(4), 363–74.
3. Coid,J. Yang, M., Tyrer, P. *et al.* (2006). Prevalence and correlates of personality disorder in Great Britain. *British Journal of Psychiatry*, **188**, 423–31.
4. Crawford, T.N., Cohen, P., Johnson, J.G., *et al.* (2005). Self-reported personality disorder in the children in the community sample: convergent and prospective validity in late adolescence and adulthood. *Journal of Personal Disorder,* **19**(1), 30–52.
5. Klein, D.N., Riso, L.P., Donaldson, S.K., *et al.* (1995). Family study of early-onset dysthymia. Mood and personality disorders in relatives of outpatients with dysthymia and episodic major depression and normal controls. *Archives of General Psychiatry*, **52**(6), 487–96.
6. Lenzenweger, M.F., Loranger, A.W., Korfine, L., *et al.* (1997). Detecting personality disorders in a non clinical population. Application of a 2-stage procedure for case identifi cation. *Archive of General Psychiatry*, **54**(4), 345–51.
7. Maier, W. (1992). Prevalence of personality disorders (DSM-III-R) in the community. *Journal of Personal Disorder*, **6**, 186–96.
8. Moldin, S.O., Rice, J.P., Erlenmeyer-Kimling, L. *et al.* (1994). Latent structure of DSM-III-R Axis II psychopathology in a normal sample. *Journal of Abnormal Psychology*, **103**(2), 259–66.
9. Reich, J., Yates, W., Nduaguba, M. (1989). Prevalence of DSM-III personality disorders in the community. *Social Psychiatry Psychiatric Epidemiology*, **24**(1), 12–6.
10. Samuels, J., Eaton, W.W., Bienvenu, O.J., III, *et al.* (2002). Prevalence and correlates of personality disorders in a community sample. *British Journal of Psychiatry*, **180**, 536–42.
11. Torgersen, S., Kringlen, E., Cramer, V. (2001). The prevalence of personality disorders in a community sample. *Archive of General Psychiatry*, **58**(6), 590–6.
12. Zimmerman, M. and Coryell, W. (1989). DSM-III personality disorder diagnoses in a nonpatient sample. Demographic correlates and comorbidity. *Archive of General Psychiatry*, **46**(8), 682–9.
13. Lenzenweger, M.F., Lane, M.C., Loranger, A.W., *et al.* (2007). DSM-IV personality disorders in the National Comorbidity Survey Replication. *Biological Psychiatry,* **62**(6), 553–64.
14. Grant, B.F., Hasin, D.S., Stinson, F.S., *et al.* (2004). Prevalence, correlates, and disability of personality disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions. *Journal of Clinical Psychiatry*, **65**(7), 948–58.
15. Baron, M., Gruen, R., Rainer, J.D., *et al.* (1985). A family study of schizophrenic and normal control probands: implications for the spectrum concept of schizophrenia. *American Journal of Psychiatry*, **142**(4), 447–55.
16. Asarnow, R.F., Nuechterlein, K.H., Fogelson, D., *et al.* (2001). Schizophrenia and schizophrenia-spectrum personality disorders in the fi rst-degree relatives of children with schizophrenia: the UCLA family study. *Archives of General Psychiatry*, **58**(6), 581–8.
17. Swartz, M., Blazer, D., George, L., *et al.* (1990). Estimating the prevalence of borderline personality disorder in the community. *Journal of Personal Disorder,* (4), 257–72.
18. Zimmerman, M. and Coryell, W.H. (1990). Diagnosing personality disorders in the community. A comparison of self-report and interview measures. *Archives of General Psychiatry*, **47**(6), 527–31.
19. Nestadt, G., Romanoski, A.J., Chahal, R., *et al.* (1990). An epidemiological study of histrionic personality disorder. *Psychological Medicine*, **20**(2), 413–22.
20. Nestadt, G., Romanoski, A.J., Brown, C.H., *et al.* (1991). DSM-III compulsive personality disorder: an epidemiological survey. *Psychological Medicine*, **21**(2), 461–71.
21. de Girolamo, G. and Tyrer, P. (1993). *Personality disorders*. Geneva: WHO.
22. Dolan, B., Evans, C., Norton, K. (1995). Multiple axis-II diagnoses of personality disorder. *British Journal of Psychiatry*, **166**(1), 107–12.
23. Zimmerman, M., Rothschild, L. and Chelminski, I. (2005). The prevalence of DSM-IV personality disorders in psychiatric outpatients. *American Journal of Psychiatry*, **162**(10),1911–8.
24. Cassin, S.E. and von Ranson, K.M. (2005). Personality and eating disorders: a decade in review. *Clinical Psychology Review*, **25**(7), 895–916.
25. Erbelding, E.J., Hutton, H.E., Zenilman, J.M., *et al.* (2004). The prevalence of psychiatric disorders in sexually transmitted disease clinic patients and their association with sexually transmitted disease risk. *Sexually Transmitted Diseases*, **31**(1), 8–12.
26. Golding, M. and Pekins, D.O. (1996). Personality disorder in HIV infection. International Review of Psychiatry, (8), 253–8.
27. Olfson, M. and Mechanic, D. (1996). Mental disorders in public, private nonprofit, and proprietary general hospitals. *American Journal of Psychiatry*, **153**(12), 1613–9.
28. Mors O. (1988). Increasing incidence of borderline states in Denmark from 1970–1985. *Acta Psychiatrica Scandinavica*, **77**(5), 575–83.
29. Loranger, A.W. (1990). The impact of DSM-III on diagnostic practice in a university hospital. A comparison of DSM-II and DSM-III in 10,914 patients. *Archives of General Psychiatry*, **47**(7), 672–5.

30. Moran, P., Jenkins, R., Tylee, A., *et al.* (2000). The prevalence of personality disorder among UK primary care attenders. *Acta Psychiatrica Scandinavica*, **102**(1), 52–77.
31. Brooke, D., Taylor, C., Gunn, J., *et al.* (1996). Point prevalence of mental disorder in unconvicted male prisoners in England and Wales. *British Medical Journal*, **313**(7071), 1524–7.
32. Singleton, N., Meltzer, H., Gatward, R., *et al.* (1998). *Psychiatric morbidity among prisoners in England and Wales*. London: Office of National Statistics.
33. Fazel, S. and Danesh, J. (2002). Serious mental disorder in 23 000 prisoners: a systematic review of 62 surveys. *Lancet*, **359**(9306), 545–50.
34. Jordan, B.K., Schlenger, W.E., Fairbank, J.A., *et al.* (1996). Prevalence of psychiatric disorders among incarcerated women. II. Convicted felons entering prison. *Archives of General Psychiatry*, **53**(6), 513–9.

4.12.5 Neuropsychological templates for abnormal personalities: from genes to biodevelopmental pathways

Adolf Tobeña

The scaffolding of personality

Research on human personality has converged upon a 'consensual pathway' indicating that a small number of dimensions can provide the framework for describing the rich variety of human temperaments. These high-level temperamental traits are factorially derived from psychometric measures of individual variation in behaviour, feeling, and thinking,[1,2] and it is assumed that they may reflect the operation of brain systems that are probably multifaceted and multipurpose.[3–5] This global outline of the structure of personality depends on the notion that genetic and developmental dispositions combine with critical nurturing and social conditioning events to form the tapestry of human uniqueness within temperamental clusters. In other words, personality types are expressed through relatively clear-cut and stable phenotypic traits that are accessible to objective measurement at behavioural, emotional, and cognitive levels. These depend, in turn, upon the specific and early organization of particular neurocognitive and neuroendocrine templates.

A handful of 'superfactors' (broad traits or dimensions) apparently capture the essential components of the mosaic of terms and traits used to describe normal personality. These dimensions are *neuroticism*, *extraversion*, *agreeableness/friendliness*, *conscientiousness*, and *intellectual openness*. Neuroticism and extraversion always appear as main stars in these factorial *solutions* whereas the remaining three superfactors—conscientiousness (reliability/persistence), friendliness (as opposed to aggressiveness/hostility) and intellectual curiosity (openness/creativity)—have less regularity on such high-order taxonomies. A five-dimensional structure is advocated by many researchers though dissent is still strong regarding the nature and scope of these superfactors that would define the 'core' of human temperament.[6]

Biological rooting of personality types

Searching for biological substrates of personality dimensions would reinforce their validity as useful constructs but this endeavour was largely neglected by psychometricians devoted to purely descriptive studies and by clinical researchers as well. Some pioneers, like Hans Eysenck at the Institute of Psychiatry, London, tried to root behavioural trait variations within neurobiological concepts[1] following a venerable tradition, which can be traced back as far as Pavlov. These early proposals were rather rudimentary though they served as drivers of subsequent models which focused more tightly on certain brain subsystems as possible sites for the factors underlying normal and abnormal temperaments.[3–5,7–9] Jeffrey Gray, Robert Cloninger, and Larry Siever's ideas were among the more fruitful in an area which has grown steadily and is now an lively field of personality research.[9–11] Progress in basic neuroscience has made it possible to relate a variety of biological measures to paper and pencil or neurocognitive tests distinguishing normal and anomalous temperaments. Biological screening has also increasingly been applied to patients with personality disorders, using the clinical clusters as defined by Diagnostic Systems. Besides these attempts to build psychobiological profiles of normal and abnormal temperaments, converging evidence is used to advocate that categorical and dimensional models for diagnosing personality disorders should be integrated.[12]

To give a broad overview of an area that may be crucial to illuminate the genesis of personality disorders, I shall discuss the studies that, during the last decade, have tried to find genetic traces for personality traits that are both behaviourally consistent and biologically well rooted. Previous work using classical (familial or twin) methods had found substantial heritability estimates for several personality traits.[13] It was thus unsurprising that genetic tracking methods impulsed research aimed at showing that temperamental traits contribute to personality scaffolding via neuroendocrine targets specified by particular genes. I'll be discussing the outcome of some of these efforts and I'll explore afterwards how other basic temperamental traits, rooted within biodevelopmental processes, do mediate enduring neurocognitive organization resulting in long-lasting behavioural styles. Finally I'll outline new avenues for the neuropsychology of personality. My approach is deliberately selective, discussing relevant evidence rather than performing a systematic assessment of the field. For reasons of convenience and possible clinical relevance, I have selected some of the traits heralding sound biological foundations, although they are not necessarily prominent in the state-of-the-art dimensional '*solutions*' for normal and abnormal temperaments.

The genetic saga for novelty-seeking

In 1996 two independent teams reported[14,15] that a particular chromosomal 'locus' was associated with a well-established trait of human temperament—the hunger for novelty and excitement that lies behind sensation-seeking, risk taking, and impulsive behaviours.[5,10] A polymorphism in the sequence of the gene expressing the D4 dopamine receptor (D4DR), located on the short arm of chromosome 11, explained 10 per cent of the genetic variance due to this trait. Individuals with the longer repeat allele at exon III of the *D4DR* gene scored higher in novelty-seeking

behaviour (explorers, risk-seekers), whereas those with the shorter allele had lower scores (prudent, cautious). The first of these studies[14] investigated a heterogeneous sample of young Israelis, and showed the association to be independent of ethnicity (Ashkenazim versus Sephardim), sex, or age. The second study, carried out in the United States,[15] used a random sample of people who had initially been recruited in a search for chromosomal regions possibly associated with sexual orientation; this sample mainly comprised white men, although some ethnic minorities were also included. The personality questionnaires were different but very popular in personality research: the Israelis were evaluated using Cloninger's Tridimensional Personality Questionnaire (**TPQ**),[16] which gives direct scores of novelty-seeking, whereas the American study used the Revised NEO Personality Inventory[17] which measures the five superfactors mentioned above, from which scores for novelty-seeking were derived. The results of the two studies were highly concordant.

Despite the modest explanatory power of this reported association, the link between temperamental variability for one trait and a chromosomal polymorphism was the first hint for a direct relationship between a putative 'genetic marker' and a core dimension of normal personality. In this case, the potential genetic marker appeared promising because of the amount of basic and clinical research linking dopaminergic function with the regulation of stimulus-seeking and sensitivity to incentives. In theory, if similar degrees of explained variance were assignable to other sound gene markers associated to *approach/exploring* phenotypes, a substantial part of the heritability of the trait could be explained. Subsequent studies[18,19] failed to replicate these early findings in a consistent way and the optimism receded. The heterogeneity of the samples and the subtleties of the genetics of complex traits were blamed for the disparate results, though the research saga was quite productive: the links between dopamine receptor polymorphisms and novelty-seeking have been intensively searched and the race to find other markers for the same trait was impressive.

Metanalyses suggest that there are subtle connections between dopamine receptor gene variants and *approach/exploring* propensities as measured by personality questionnaires, though the strength of the contribution of every variant is small and hard to establish.[19,20] Moreover, parallel research has established suggestive connections between gene variants regulating other molecular targets (i.e. tryptophan hydroxylase, dopamine transporter, dopamine-beta-hydroxylase, serotonin transporter, MAOA, COMT) that modulate risk-taking behaviours and impulsivity. A handful of genes, thus contribute to differential vulnerabilities for addictive behaviours, a congruent result at the extreme of stimulus-seeking tendencies. Although further and more refined research is required, these data seem to confirm pioneer work, mostly with twins, which had consistently established that novelty-seeking behaviour was moderately heritable (40 to 50 per cent).

Genetics of fearfulness/neuroticism

The aforementioned American team that reported the first associations between novelty-seeking and variants of D4DR gene informed that there was an association between the neuroticism trait and a chromosomal region linked to serotonin neurotransmission involved in modulating anxiety-related traits.[21] The 5-hydroxytryptamine transporter protein (5-HTT) that promotes the reuptake of serotonin into cell membranes is encoded by a gene (*SLC6A4*) located in the q11–q12 segment of chromosome 17. The region governing the transcriptional control of the protein shows a polymorphism that influences its expression and functioning. Individuals carrying the short variant of the polymorphism show a reduced efficiency of serotonin reuptake compared with those possessing the longer variant. The study measured these parameters in the lymphoblasts of two independent samples totalling more than 500 volunteers. Using two different personality questionnaires (NEO and Cattell's 16PF Personality Inventory) and estimated scores on various dimensions of Cloninger's TPQ, the evidence showed that subjects who carried the short variant in the 5-HTT gene polymorphism had higher neuroticism (NEO), anxiety (16PF), and harm-avoidance (TPQ) scores. The results were equally consistent across and within pedigrees. Across the three personality measures, the 5-HTTLPR contributed a modest 3 to 4 per cent of the total variance and 7 to 9 per cent of the genetic variance in anxiety-related traits. It was suggested that, if other genes contributed similar dosage effects to anxiety traits, approximately 10 to 15 genes might be involved in the heritability of neuroticism.

The implication of the serotonin transporter in the potential genetic predisposition towards emotionality traits agrees with many other results. Serotonergic neurotransmission is involved in multiple brain functions with little or no relationship with fear/anxiety regulation, but there is a large body of evidence linking it with the modulation of adaptive responses to serious conflict and emotionally demanding situations.[3] Moreover, many drugs currently used to treat anxious/depressive dysphorias and personality disorders depend on mending serotonergic function. Finally, several studies have shown that variants of the 5-HTTLPR predict differential response on anxious phenotypes: fear-driven amygdala activation[22] and response to pharmacological challenges.[23] Therefore, although the exact role of serotonergic systems in the modulation of emotionality is not fully understood, it is improbable that the neurohormonal adaptations that participate in individual responses to serious emotional conflicts would not include serotonin modulation either through the cell transporter or through the extended family of serotonin receptors and their intracellular targets. Defensive adaptations require however the participation of other central neuromodulators: the CRH-ACTH regulatory cascade, γ-aminobutyric acid, neuropeptide Y, and substance P,[3] are major contenders in this respect and they can be expected to contribute to the genetic mediation of neuroticism.

Polymorphisms in anxious humans versus QTL-genes for fearful rodents

Dozens of studies investigated whether the particular polymorphism in the 5-HTT gene contributes to the tendency for individuals who score higher on neuroticism, in personality tests, to be at higher risk for 'internalizing disorders' (anxiety/depression) and personality disorders (anxiety/affective clusters). The global outcome of that research has been unreliable: though stringent metanalyses confirmed the original association with a modest relevance at explaining the trait variance,[23] subsequent results in large samples of siblings and singletons have been negative.[24] There are other lines of evidence, however, mainly from animal research, that support the claim of a possible genetic basis for fearfulness/emotionality. This evidence is derived from studies of the

psychogenetics of emotional susceptibility searching for chromosome loci. In many biological and behavioural tests, comparisons of several reactive and non-reactive strains of mice and rats obtained through artificial selection (forced interbreedings) have narrowed the search for genetic loci thanks to increasingly powerful methods of chromosomal mapping. In a pioneer work with progeny obtained by crossing two strains of mice selected for activity and defecation in an open-field test, three loci (QTLs) which explained most of the genetic variance in emotionality were found on chromosomes 1, 12, and 15 of the murine genome.[25] These data were confirmed and extended by measuring fear responses towards particular cues: the same segment of chromosome 1 was identified as a relevant 'locus' for emotional susceptibility besides other murine chromosomal zones.[9] The importance of the loci at chromosome 1 has been established in studies using heterogeneous and inbred stocks, combining techniques which have permitted to focus the suspicious segment to less than 1 cM and leading to the identification of the first gene linked to murine emotionality: Rgs2, a regulator of G-protein functioning which is highly expressed in the brain.[26] The complexities are nevertheless tremendous because even a QTL like that contains several genes each contributing a very modest part of variation on the phenotype of interest.[9,26]

Several research programmes were concurrently started to determine whether there is also concordance in the chromosomal marking of emotionality in strains of rats differing in fearfulness. The hypoemotional Roman high-avoider (**RHA**) versus the hyperemotional Roman low-avoider (**RLA**) rat lines represented a particularly interesting assay because of the very large body of evidence showing their usefulness as animal models of 'temperamental' styles.[27] Several plausible QTLs were detected but only those located on chromosome 5 and 15 predicted a wide array of anxious/fearful behaviours including spontaneous and learned fears.[28,29] The QTL located at the middle of rat chromosome 5 looks particularly promising because this region is partially syntenic with the human 1p segment where QTLs for neuroticism and liabilities for anxious/depressive dysphorias have been detected.[9,30] The search for plausible genes is, then, much more focused now though they will explain, in all probability, very modest portions of the phenotypic trait variance.[9]

Biodevelopmental mechanisms for affiliative traits

Affiliation (friendliness, sociability, gregariousness, empathy) is another core personality trait that can be measured consistently. Poor or distorted affiliative behaviour is the most predictive marker for a reliable diagnosis of personality disorder.[7,31] Deficits or alterations in affiliative tendencies may show a variety of clinical manifestations: extreme aloofness and detachment, manipulative, non-empathic or exploitative attitudes, and even definite asocial or antisocial tendencies. These behavioural styles appear, in different degrees and combinations, in several clinical categories of abnormal temperament. They could reflect alterations in the functioning of neuroendocrine systems specialized in mediating affective attachments, possibly including subsystems for social reward and social distress.[31] In this respect, an impressive amount of evidence has been gathered (mostly in animals but in humans as well) showing that prosocial behaviours, such as maternal nurturing, friendliness/gregariousness, playful/sexual behaviours, and even cooperativeness in economic interactions, share some neurochemical controls.[31–33]

Central oxytocin and opioid systems are among the more relevant of these modulatory molecules, because several types of attachments (mother–infant bonds, young peer bonds, pair-mating bonds, in-group tendencies) are dramatically altered when the functions of oxytocinergic or opioid systems in the brain are modified. Other centrally acting neuropeptides, such as prolactin and vasopressin, also contribute to particular types of species-specific social bonding. In addition, both the central regulatory monoaminergic systems and the corticotrophin–adrenal cascades that mediate stress adaptations help to organize responses to everyday social challenges.[31] The application of these ideas to personality is still in its infancy and requires the development of consistent scales to be related with sound biological markers.

Neuropeptides, social bonding, and early rearing practices

Theory and research in the psychobiology of social attachments[8,31] has linked the impact of early rearing practices (secure/nurturing mothering versus peer rearing or isolation) with the future organization and functioning of several neuroendocrinological systems. This research has mainly explored the function of the central modulatory monoamines norepinephrine, dopamine, and serotonin, and the hypophyseal–adrenal axis responses to social challenges. The evidence has shown that socially deprived monkeys differ physiologically, behaviourally, and cognitively from mother-reared infants in almost every aspect of what it means to be a social monkey; if the privation extends over the first 6 to 9 months of post-natal life, most effects persist into the adulthood. According to Kraemer[34]: 'The way in which socially deprived individuals orient to and respond physiologically to social stressors is altered and the kind of behavioural differences that seem to be the most important are those usually assigned to the domain of "temperament"'.

With the addition of the central neuropeptide systems that specifically modulate affiliative tendencies, the study of some crucial experiences during early infancy (and probably adolescence) will provide clues to the clarification of the role that developmental processes play in shaping attachment styles. Neural organization depends, to a great extent, on critical environmental inputs to produce enduring behavioural and cognitive adaptation in all domains. Therefore ontogenic factors must be particularly important in modelling social behaviour and in sustaining profiles of affiliative versus non-affiliative temperaments, in the same way as has already been demonstrated for other traits. For instance, in reactive/fearful monkeys, maternal and even grandparental rearing practices can significantly modify future neuroendocrine and behavioural adaptations[35] with parallel data obtained in rats.[36]

Affiliative genes?

These findings do not exclude the participation of genetic dispositions in attachment styles. Some authors have suggested that the operation of 'communicative' or 'affiliative' genes could prime individual tendencies through different sorts of emotional affects.[37] There is a paucity of data on the putative genetic basis of particular attachment styles. When a molecular approach has been used,

in rodents, to establish a genetic link between variants of the vasopressin V1a receptor gene with a conspicuous social behaviour such as monogamous pair bonding, the results have been spectacular. After substituting and inserting the V1a receptor gene characteristic of a monogamous species (the prairie voles), both promiscuous mice and meadow voles adopted the pattern of partner preferences and parental behaviours distinctive of the monogamous species. It has been shown, in fact, that socio-behavioural trait differences depend on polymorphisms on the regulatory region of the V1a gene.[38] In rodents and other mammals the neurochemical circuits in the brain regulating attachment/affiliative behaviours are increasingly detailed. In humans the task is just starting and will deeply influence personality research.

Biodevelopmental processes for aggressiveness

Aggressiveness is another temperamental trait that has a very well-founded biodevelopmental basis.[39] Dimensional descriptions of the structure of personality do not always include aggressiveness as a high-level factor, but it is embedded in other traits such as impulsiveness, poor control, or explosive/desinhibited behaviour. However, aggressiveness is a major behavioural trait that has to be taken into account in the routine management of mental disorders, and it is also a prevalent characteristic in personality disorders (generally as an excess, but sometimes because of its absence). In the past, it was extremely difficult to prove that individual variations in aggressive behaviour might be correlated with neurohormonal characteristics. However, this was because of inadequate methods of quantifying biological and behavioural variations.[40]

In humans, the link between lifelong aggressive profiles and familial/subcultural problems is strong indicating that lower socio-economic status, increased rates of abuse, coercive family interactions, and neglect or other adversities contribute to violent behaviour from infancy to adolescence and into adulthood.[39,41] But this cannot obscure the effects of enduring biological dispositions that could, in part, result from the influence of socio-environmental insults to the developing brain. Intensive research in behavioural neuroendocrinology and molecular neuroscience using animal models has shown that aggressive temperaments are associated with both genetic dispositions and critical developmental processes, which affect the function of neuroregulatory controls that either promote or inhibit aggression.

MAO-A functioning and aggression-proneness

It was not surprising, therefore, that an association between a specific gene and an aggressive disorder in humans was early reported[42]: the males of a Dutch family that carried a mutation of the gene for monoaminoxidase A (MAO-A gene situated at chromosome Xp11.23) had a record of severe aggressive incidents in different generations. Subsequent investigations in mice showed that ablating the gene for MAO-A resulted in the 'knockout' mutants being much more aggressive[43] MAO plays an important part in the breakdown of serotonin and other central monoamines, and previous research in humans had found consistent relationships between impulsive or 'poorly controlled' behavioural styles and low-MAO activity, a feature also seen in sensation-seekers. A common polymorphism in the promoter region of the MAO-A gene has been shown to interact with early abuse/neglect in longitudinal studies of large samples of children[44]: only abused kids carrying the genetic variant giving low-MAO-A activity are later at risk of antisocial tendencies or violent/criminal behaviour, during adolescence or young adulthood. Further studies with normative samples differentiated by this polymorphism have permitted to map, through structural and functional neuroimaging, cortico-limbic singularities associated with emotional regulation in the low-MAO-A carriers.[45]

Serotonin/vasopressin ratios

Many other genetically altered animals have been produced that are highly aggressive. Several of them typically show anomalies in serotonin function. Knockout mice that do not express the serotonin 5-HT1B receptors are much quicker to attack intruders[46] an action that can be blocked with targeted serotonergic drugs. All this adds to the evidence obtained from mentally ill patients presenting aggressive outbursts and from chronic offenders in prisons, which shows an inverse relationship between serotonergic function and violent attacks directed towards others or themselves. Taken together, the data appear to suggest that preserved brain serotonin function helps to attenuate aggressive impulses, probably by blocking other neuroregulatory systems that promote aggression such as sex steroids, insulin, vasopressin, and others. Vasopressin is involved in the establishment of pair bonds and territorial tendencies in mammals, and promotes aggressive behaviour when social/dominance challenges are perceived. There is evidence from adolescent and juvenile hamsters showing that proserotonergic interventions attenuate vasopressin-induced attacks. Patients with personality disorders[47] showed that cerebrospinal fluid levels of vasopressin were positively correlated with a life history of aggression, and with attacks against persons in particular. This was a more poweful relationship than the negative one usually obtained from measurements of serotonergic function and aggression. Other neuromodulators help serotonin to attenuate aggressive outputs: in animals, highly specific genetic techniques of knocking down or targeted regional brain expression of steroid and oxytocin receptors, have been used to identify ensambles of neuromodulators devoted to control aggressiveness, to the point of postulating a landscape of 50 genes to get a sound description to this function.[48]

A prospective landscape including other traits

Temperamental styles in animals show differential clustering of behavioural traits, which correspond to specific (and very complex) neurohormonal profiles. Sometimes, however, a specific genetic modification is sufficient to promote a fully differentiated temperament, as in the case of mice deficient in α-calcium–calmodulin-dependent kinase II (**α-CaMKII**) which show decreased fear responses and increased defensive aggression associated with low-serotonin levels[49] (Table 4.12.5.1).

It is possible that this may also hold for exceptional temperamental combinations in humans. However, we have already established that, when trying to explain the genetic contribution to basic ('universal') personality traits, a multigenic/interactional approach is compulsory to explain just part of the measured variance in each trait (Fig. 4.12.5.1). Nevertheless, this contribution can be very

Table 4.12.5.1 Summary of behavioural phenotypes in α-CaMKII mutant mice

Behavioural phenotype	Heterozygote	Homozygote
Fear-related responses	Decrease	Decrease
Offensive aggression	Normal	Decrease
Defensive aggression	Increase	Decrease
Pain sensitivity	Normal	Increase
Startle response	Normal	Increase
Vigilance	Normal	Increase
Mating	Decrease	Decrease
Maze learning	Normal	Decrease

Reproduced from C. Chen *et al.* Abnormal fear response and aggressive behaviour in mice deficient for alpha–calcium–calmodulin–leinase II. *Science*, **266**, 291–4, copyright 1994, with permission from the American Association for the Advancement of Science.

important, as most of research involving twins has yielded estimates of just over 40 per cent for the genetic input to typical personality traits, with very modest estimations assignable to the so-called shared (familial/cultural) environment. Non-shared environmental influences (from the womb onwards) and genetic–environmental interactions make a well-known contribution to each individual temperament. These complex influences may act first by modelling the development of basic neuroendocrine regulatory systems that cope with natural and social challenges, and second by shaping the neurocognitive architecture that results in an autonomous and particular lifestyle. An interactional approach along these lines is now laying the foundation for a better description of the different factors that contribute to building the typical profiles of normal or abnormal personalities.[50,51]

Such a general scheme must include other traits, in addition to the ones considered so far. For instance, the detection of a substantial genetic contribution to the baseline level of happiness,[52] which all individuals show throughout life independently of the events or episodes that they encounter, must be important for

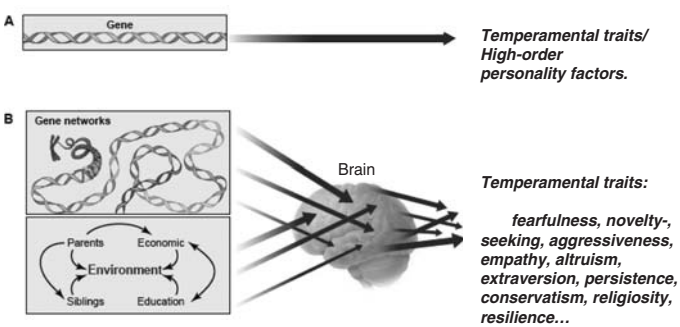


Fig. 4.12.5.1 Two views of the relationships between genes and personality. **A**. Early studies looked for linear relationships between gene markers and temperamental traits or high-order personality factors. **B**. Reality is likely to be far more complex with gene networks interacting with environmental inputs impacting on brain development and leaving enduring neural dispositions that, in turn, influence behavioural, cognitive, and affective styles (temperamental traits). (Reproduced from D. Harner. Rethinking behavior genetics, *Science*, **298**, 71–2, copyright 2004, with permission from the American Association for the Advancement of Science.).

personality diagnoses, because such a 'calibration point' has a direct relationship with the affective tone of optimism or pessimism. Moreover, other traits and measures in the domain of cognitive performance (e.g. attentional spans, perceptual/appraisal reliability, thinking styles, and memory biases) or of character (e.g. religiosity/transcendence, conservatism, altruism, self-directness/self-esteem, and the drive to achieve/enthusiasm) should be incorporated into the whole description of personality structure. This is essential if the aim is to produce a general framework powerful enough to contain the complex categories that clinicians have tried to construct on the basis of systematic observation for more than a century. Cloninger[8,11] and others have advanced proposals along these lines, which have functioned as useful steps to guide empirical work. It must be added that structural and functional neuroimaging techniques have been used to map relations between personality dimensions and brain's regional organization and patterns of activation/deactivation. Although these studies are very preliminary and have been done with convenient samples, they have offered hints of neurocognitive architectures, which tend to confirm, on a broad basis, the neurochemical mechanisms behind the main temperamental traits.[53–55]

Sophisticating the diagnoses of personality disorders

The evidence discussed so far is starting to have a major impact on the reconceptualizing of the approach that has been applied in psychiatry to the detection and categorization of the elusive profiles of normal and abnormal personalities. It is clear that anchoring some temperamental traits to a sound genetic or biodevelopmental base, and accruing a plausible neurocognitive architecture for it, will not provide a complete solution to the problems encountered in the aetiology and diagnosis of personality disorders. Many additional steps will have to be worked out. But in view of the increasing links that are being established between core personality traits and some genetic/developmental mechanisms, the task of building a solid neuropsychological framework, which would allow improved identification and differentiation of abnormal personalities no longer seems hopeless. In fact, such research is opening up many new avenues for understanding the effects of different factors (innate dispositions, neurodevelopmental organization, neurocognitive architectures, critical social transitions, and repeated stress episodes) on an individual's vulnerability to developing a personality disorder.

Hence, the use of behaviourally well defined and biologically well-established personality measures must be the starting point for achieving the fine-tuned diagnoses increasingly required in modern medicine. Complexity in measurement will increase, but there is no other way of obtaining data that are sufficiently valid to allow understanding of the classical personality disorders. It is hoped that advances in the neuropsychological detection of the more salient 'clinical' profiles within each of the personality subspaces or 'clusters' (at the level of either traits or dimensions), together with a refinement of the neurocognitive and neurohormonal data, will produce much better solutions. Some of the old-established and consistent categories of personality disorders will be confirmed, but it is possible that unexpected 'types' of abnormal personalities, with clinical relevance, will emerge. In this new

framework it may be easier perhaps to detect at an early stage those 'exceptional' and 'charismatic', although anomalous, personalities who often impose great social costs and dramatic consequences not only for themselves but also for the group or the society in which they live.[56]

Further information

Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neuroscience and Biobehavioral Reviews*, **30**, 319–45.

Caspi, A., Roberts, B.W., and Shiner, R.L. (2005). Personality development: stability and change. *Annual Review of Psychology*, **58**, 453–84.

Gosling, S.D. (2001). From mice to men: what can we learn about personality from animal research. *Psychological Bulletin*, **127**, 45–86.

Hariri, A.R. and Holmes, A. (2006). Genetics of emotional regulation: the role of serotonin transporter in neural function. *Trends in Cognitive Sciences*, **10**, 182–91.

Kreek, M.J., Nielsen, D.A., Butelman, E.R., *et al.* (2005). Genetic influences on impulsivity, risk taking, stress responsivity vulnerability to drug abuse and addiction. *Nature Neuroscience*, **8**, 1450–7.

Markon, K.E., Krueger, R.F., Bouchard, T.J. Jr, *et al.* (2002). Normal and abnormal personality traits: evidence for genetic and environmental relationships in the Minnesota study of twins reared apart. *Journal of Personality*, **70**, 660–93.

McGue, M. and Bouchard, T.J. (1998). Genetic and environmental influences on human behavioral differences. *Annual Review of Neuroscience*, **21**, 1–24.

McNaughton, N. and Corr, P..J. (2004). A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neuroscience and Biobehavioural Reviews*, **28**, 285–305.

Paris, J. (2005). Neurobiological dimensional models of personality: a review of the models of Cloninger, Depue and Siever. *Journal of Personality Research*, **19**, 156–70.

Tobeña, A. (2004). *Deadly martyrs: the biology of lethal altruism*. Bromera Publish.-Valencia University Press, Valencia, Spain.

Young, L.J. and Wang, Z. (2005). The neurobiology of pair bonding. *Nature Neuroscience*, **7**, 1048–54.

References

1. Eysenck, H.J. (1967). *The biological basis of personality*. Charles Thomas, London.
2. McRae, R.R. and Costa, P.T. (1997). Personality trait structure as a human universal. *The American Psychologist*, **52**, 509–16.
3. Gray, J.A. and McNaugthon, N. (2000). *Anxiety: an enquiry into the functions of the septohippocampal system* (2nd edn). Oxford University Press, New York.
4. Cloninger, C.R. (1987). A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry*, **44**, 573–88.
5. Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. Cambridge University Press, New York.
6. Sheller, J. and Westen, D. (2004). Dimensions of personality pathology: an alternative to the five-factor model. *The American Journal of Psychiatry*, **161**, 1743–54.
7. Siever, L.J. and Davis, K.L. (1991). A psychobiological perspective on the personality disorders. *The American Journal of Psychiatry*, **148**, 1647–58.
8. Cloninger, C.R., Svrakic, D.M., and Przybeck, T.R. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry*, **50**, 975–90.
9. Flint, J. (2004). The genetic basis of neuroticism. *Neuroscience and Biobehavioural Reviews*, **28**, 307–16.
10. Corr, P.J. (2004). Reinforcement sensitivity theory and personality. *Neuroscience and Biobehavioural Reviews*, **28**, 317–32.
11. Sravic, D.M., Draganic, S., Hill, K., *et al.* (2002). Temperament, character and personality disorders: ethiogic, diagnostic, treatment issues. *Acta Psychiatrica Scandinavica*, **106**, 189–95.
12. Widiger, T.A. and Frances, A.J. (2001). Towards a dimensional model for personality disorders. In *Personality disorders and the five-factor model of personality* (2nd edn) (eds. P.T. Costa and T.A. Widiger). American Psychological Association Press, Washington, DC.
13. Markon, K.E., Krueger, R.F., Bouchard, T. J. Jr, *et al.* (2002). Normal and abnormal personality traits: evidence for genetic and environmental relationships in the Minnesota study of twins reared apart. *Journal of Personality*, **70**, 660–93
14. Ebstein, E.P., Novick, O., Umansky, R., *et al.* (1996). Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of novelty seeking. *Nature Genetics*, **12**, 78–80.
15. Benjamin, J., Li, L., Greenberg, B.D., *et al.* (1996). Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nature Genetics*, **12**, 81–4.
16. Cloninger, C.R., Przybeck, T.R., and Svrakic, D.M. (1991). The tridimensional personality questionnaire: US normative data. *Psychological Reports*, **69**, 1047–57.
17. Costa, P.T. and McCrae, R.R. (1992). *Revised NEO personality inventory and NEO five inventory*. Psychological Assessment Resources, Odessa, FL.
18. Lusher, J.M., Chandler, C., and Ball, D. (2001). Dopamine D4 receptor gene (DRD4) is associated with novelty seeking (NS) and substance abuse: the saga continues. *Molecular Psychiatry*, **6**, 497–9.
19. Shinka, J.A., Letsch, E.H., and Crafword, F.C. (2002). DRD4 and novelty seeking: results of a metanalyses. *American Journal of Medical Genetics*, **44**, 643–8.
20. Munafo, M.R., Clark, T.G., Moore, L.R., *et al.* (2003). Genetic polymorphisms and personality in healthy adults: a systematic review and metanalyses. *Molecular Psychiatry*, **8**, 471–84.
21. Lesch, K.P., Bengel, D., Heils, A., *et al.* (1996). Association of anxiety related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, **274**, 1527–31.
22. Hariri, A.R., Mattay, V.S., Tessitore, A., *et al.* (2002). Serotonin transporter genetic variation and the response of the human amygdala. *Science*, **297**, 400–03.
23. Smith, G.S., Lothich, F.E., Molhutra, A.K., *et al.* (2004). Effects of serotonin transporter promoter polymorphisms on serotonin function. *Neuropsychopharmacology*, **29**, 2226–34.
24. Willis-Owen, S.A.G., Turri, M.G., Munafò, M.R., *et al.* (2005). The serotonin transporter length polymorphism, neuroticism and depression: a comprehensive assessment of association. *Biological Psychiatry*, **58**, 451–6.
25. Flint, J., Carley, R., DeFries, J.C., *et al.* (1995). A simple genetic basis for a complex psychological trait in laboratory mice. *Science*, **268**, 1432–5.
26. Yakin, B., Willis-Owen, S.A.G., Fullerton, J., *et al.* (2004). Genetic dissection of a behavioral quantitative trait locus shows the Rgs2 modulates anxiety in mice. *Nature Genetics*, **36**, 1197–202.
27. Driscoll, P., Escorihuela, R.M., Fernández-Teruel, A., *et al.* (1998). Genetic selection and differential stress responses: the Roman lines/strains of rats. *Annals of the New York Academy of Sciences*, **851**, 521–30.
28. Fernandez-Teruel, A., Escorihuela, R.M., Gray, J.A., *et al.* (2002). A quantitative trait locus influencing anxiety in the laboratory rat. *Genome Research*, **12**, 618–26.
29. Aguilar, R., Gil, L., Flint, J., *et al.* (2002). Learned fear, emotional reactivity and fear of heights: a factor analytic map from a large F2 intercross of Roman rat strains. *Brain Research Bulletin*, **57**, 17–26.
30. Nash, M.W., Huezo Diaz, P., Williamson, R.J., *et al.* (2004). Genome-wide linkage analyses of a composite index of neuroticism and mood related scales in extreme selected sibships. *Human Molecular Genetics*, **13**, 2173–82.

31. Panksepp, J. (1998). *Affective neuroscience*. Oxford University Press, New York.
32. Insel, T.R., Young, L., and Wang, Z. (1997). Molecular aspects of monogamy. *Annals of the New York Academy of Sciences*, **807**, 302–16.
33. Kosfeld, M., Heinrichs, M., Zak, P.J., *et al.* (2005). Oxytocine increases trust in humans. *Nature*, **435**, 673–6.
34. Kraemer, G.W. (1997). Psychobiology of early social attachment in rhesus monkeys. In *The integrative neurobiology of affiliation* (eds. C.S. Carter, I.I. Lederhendler, and B. Kirpatrick). *Annals of the New York Academy of Sciences*, **807**, 401–18.
35. Suomi, S.J. (1991). Uptight and laid-back monkeys: individual differences in the response to social challenges. In *Plasticity and development* (eds. S.E. Brauth, W.S. Hall, and R.J. Dooling), pp. 27–56. MIT Press, Cambridge, MA.
36. Liu, D., Diorio, J., Tannenbaum, B., *et al.* (1997). Maternal care, hippocampal glucocorticoid receptors and hypothalamic-pituitary-adrenal responses to stress. *Science*, **277**, 1659–62.
37. Buck, R. and Ginsburgh, B. (1997). Communicative genes and the evolution of empathy: selfish and social emotions as voices of selfish and social genes. *Annals of the New York Academy of Sciences*, **807**, 481–3.
38. Hammock, E.A.D. and Young, L.J. (2005). Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science*, **308**, 1630–4.
39. Tremblay, R.E., Hartup, W.W., and Archer, J. (eds.) (2005). *Developmental origins of aggression*, Guilford Books, New York.
40. Adams, D.B. (2006). Brain mechanisms of aggressive behavior: an updated review. *Neuroscience and Biobehavioral Reviews*, **30**, 304–18.
41. Ferris, C.F. and Grisso, T. (eds.) (1996). Understanding aggressive behavior in children. *Annals of the New York Academy of Sciences*, **794**, 98–103.
42. Brunner, H.G., Nelen, M., Breakefiled, X.O., *et al.* (1994). Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. *Science*, **262**, 578–80.
43. Cases, O., Seif, I., Gromsby, J., *et al.* (1995). Aggressive behavior and altered amounts of brain serotonin and noradrenaline in mice lacking MAO-A. *Science*, **268**, 1763–6.
44. Caspi, A., McClay, J., Moffit, T., *et al.* (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, **297**, 851–4.
45. Meyer-Lindenberg, A., Buckholtz, J.W., Kolachana, B., *et al.* (2006). Neural mechanisms of genetic risk for impulsivity and violence in humans. *PNAS*, **103**, 6269–74.
46. Saudou, F., Amarada, G., Dierich, A., *et al.* (1994). Enhanced aggressive behavior in mice lacking 5-HT1b receptor. *Science*, **265**, 1875–8.
47. Coccaro, E.F., Kavoussi, R.J., Hauger, R.L., *et al.* (1998). Cerebrospinal fluid vasopressin levels: correlates with aggression and serotonin function in personality-disordered patients. *Archives of General Psychiatry*, **55**, 708–14.
48. Ogawa, S., Choleris, E., and Pfaff, D. (2004). Genetic influences on aggressive behaviors and arousability in animals. *Annals of the New York Academy of Sciences*, **1036**, 257–66.
49. Chen, C., Rainnie, D.G., Greene, R.W., *et al.* (1994). Abnormal fear response and aggressive behavior in mice deficient for alpha-calcium-calmodulin-kinase II. *Science*, **266**, 291–4.
50. Saudino, K.J. (2005). Behavioral genetics and child temperament. *Journal of Developmental and Behavioral Pediatrics*, **26**, 214–23.
51. Blackburn, R., Logan, C., Renwick, S.D., *et al.* (2005). Higher-order dimensions of personality disorder: hierarchical structure and relationships with the five factor model, the interpersonal cycle and psychopathy. *Journal of Personality Disorders*, **19**, 597–623.
52. Lykken, D.T. and Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science*, **7**, 186–9.
53. Pujol, J., Lopez, A., Deus, J., *et al.* (2002). Anatomical variability in the anterior cingulate gyrus and basic dimensions of personality, *Neuroimage*, **15**, 847–55.
54. Kumari, V., Ffytche, D.H., Williams, S.C.R., *et al.* (2004). Personality predicts brain responses to cognitive demands. *The Journal of Neuroscience*, **24**, 10636–41.
55. O'Gorman, R.L., Kumaro, V., Williams, S.C.R., *et al.* (2006). Personality factors correlate with regional cerebral perfusion. *Neuroimage*, **31**, 489–95.
56. Henry, D., Geary, D., and Tyrer, P. (1993). Adolf Hitler: a re-assessment of his personality status. *Irish Journal of Psychological Medicine*, **10**, 148–51.

4.12.6 Psychotherapy for personality disorder

Anthony W. Bateman and Peter Fonagy

Introduction

Psychotherapy has historically been the mainstay of treatment for personality disorder (PD). It remains so. Psychoanalysis was probably the earliest formal treatment for PD, which led to the first clinical descriptions of borderline personality disorder. A parallel but linked development was the application of psychoanalytic ideas in therapeutic communities which have been in existence for over 60 years and remain a treatment context and method for patients with PD. It was only in the 1960s that modified psychotherapeutic treatments were developed. Initially these were based on psychodynamic understanding of PD, but gradually other theoretically and practically driven models have developed, leading to the current situation in which there are behavioural, cognitive, dynamic, and supportive treatments offered in a range of contexts. Some of these methods have more empirical support than others. These methods will be described in this chapter.

Psychological therapies for personality disorders take place against the background of the natural course and outcome of the disorder. Until recently, the natural history of personality disorder had not been systematically studied. Several major cohort follow-along studies have yielded surprising data concerning the rate of symptomatic remissions in a disorder that was assumed to have a lifelong course.[1] For example, over a 10-year follow-along period, 88 per cent of those initially diagnosed with borderline personality disorder appeared to remit in the sense of no longer meeting DIB-R or DSM-III criteria for BPD for 2 years.[2] The symptoms that remit most readily, irrespective of treatment, appear to be the acute ones, such as parasuicide and self-injury, which are the most likely to trigger psychotherapeutic intervention. Temperamental symptoms, such as angry feelings and acts, distrust and suspicion, abandonment concerns, and emotional instability, appear to resolve far more slowly. In the Collaborative Longitudinal Personality Disorder Study (CLPS),[3] when remission was defined as 12 months at two or fewer criteria for PDs, over half of BPD and 85 per cent of major depressive disorder (MDD) patients were reported to remit over a 4-year period. Psychosocial functioning recovered far more slowly than acute symptoms.[1]

There is a considerable body of literature on psychotherapeutic interventions for personality disorders, but significant evidence

for effective treatment remains sparse. Much of the literature is dominated by expert opinion, which is not invariably the most helpful guide. In this chapter, we focus on psychological treatments where at least some evidence for treatment effectiveness exists. The evidence is strongest for borderline personality disorder (BPD). Treatment of some other personality disorders, for example schizoid, narcissistic, obsessive–compulsive, dependent, is evidenced mainly by clinical case reports in which theory is combined with clinical description and where, if outcome is measured at all, it is measured for the purpose of illustration and has little probative value.

Assessment of treatment

Any study that seeks to demonstrate the effectiveness of a treatment for PD must fulfil the following requirements:

(a) *Carefully define the target population*. This can be problematic, because the definition of personality disorder (PD) remains controversial, and there is little evidence that the categories of personality disorder have any predictive value in determining response to treatment. Comorbidity must also be considered. Lifetime comorbidity should not be an exclusion criterion for studies, but there is an argument for excluding individuals with current comorbidity. However, in clinical practice, such exclusion is almost impossible.

(b) *Adequately define the treatment and assessment of its specificity*. Personality disorder is a multifaceted condition that is susceptible to a variety of influences, and it justifies the use of complex interventions. These require complex evaluations, which increase the difficulty of interpreting results.

(c) *Establish that treatment is superior to no treatment* since personality disorders show gradual improvement over time.

(d) *Take account of Axis I disorders*. This can be done by excluding patients with Axis I disorders, but PDs are almost always associated with significant Axis I psychopathology. An alternative, which no trial has, to date, attempted, would be to assign patients to treatment group on the basis of matched Axis I disorders. In any case, it must be demonstrated that treatment impacts on personality rather than merely causing a change in mood or psychiatric symptoms.

(e) *Include adequate follow-up*, as some trials report reduced treatment effects during follow-up.

(f) *Address cost-effectiveness* relative to alternative interventions.

(g) *Study treatment effects in standard practice* (pragmatic trials) as well as under strict experimental conditions.

Research trials investigating the effectiveness of treatments for personality disorder have so far singularly failed to meet most of these requirements.

Adverse effects of psychotherapy for personality disorder

It is possible that some psychosocial treatments for personality disorder may have impeded the patient's capacity to recover following the natural course of the disorder and/or prevented them from taking advantage of changes in social circumstances. In Stone's[4] classic follow-up of patients treated nearly 40 years ago, a 66 per cent recovery rate was only achieved in 20 years, about four times longer than reported in the more recent studies. It seems unlikely that the nature of the disorder has changed or that treatments have become markedly more effective rather it is possible that treatments with these adverse effects are being offered less frequently now than in the past, perhaps because of changed patterns of health care, particularly in the United States.

Meta-analyses of psychotherapy and psychosocial treatments

It remains unclear whether the literature is robust enough to withstand the methodology of meta-analysis. The lack of good quality studies, especially randomized trials, the small number of patients in the trials, the heterogeneity of the personality disorders studied, and the variability of outcome measures across studies means that conclusions must remain tentative. One meta-analysis[5] included 15 studies that reported data on pre- to post-treatment effects and/or recovery at follow-up, including three randomized, controlled trials, three randomized comparisons of active treatments, and nine uncontrolled observational studies. They included psychodynamic/interpersonal, cognitive behavioural, mixed, and supportive therapies. All studies reported improvement in personality disorders with treatment. The mean pre–post effect sizes within treatments were 1.11 for self-report measures and 1.29 for observational measures. Among the three randomized, controlled treatment trials, active psychotherapy was more effective than no treatment according to self-report (ES = 0.75), though none of the controls employed an active therapy. Only four studies reported the percentage of cases no longer meeting the criteria for personality disorder. At follow-up (at a mean of 67 weeks), 52 per cent met this criterion. Treatment length was associated with the likelihood of recovery.

A subsequent meta-analysis[6] included psychodynamic therapy and cognitive behaviour therapy (CBT) in the treatment of personality disorders. There were 22 studies of psychodynamic therapy and CBT published between 1974 and 2001 that (1) used standardized methods of diagnosis, (2) applied reliable and valid instruments for the assessment of outcome, and (3) reported data that allowed calculation of within-group effect sizes or included assessment of recovery rates. Because only 11 of the 22 studies were RCTs, ESs were calculated on the basis of pre- to post-therapy change. Fourteen studies included psychodynamic therapy, and 11 studies included CBT. The psychodynamic studies had a mean follow-up of 1.5 years, compared to only 13 weeks for CBT. Psychodynamic therapy yielded an overall effect size of 1.46 ($k = 15$ contrasts), with effect sizes of 1.08 for self-report measures and 1.79 for observer-rated measures. For CBT ($k = 10$ contrasts), the corresponding values were 1.00, 1.20, and 0.87. For more specific measures of personality disorder pathology, a large overall effect size (1.56) for psychodynamic therapy suggested long-term rather than short-term change in personality disorders. For BPD, the ES for psychodynamic therapy was 1.31 ($N = 8$), and for CBT 0.95 ($N = 4$). Treatment length showed a positive but nonsignificant correlation with outcome in psychodynamic studies ($r = 0.41$); there were too few CBT trials for an equivalent analysis. In the 5 years since 2001, there have been nearly as many new randomized trials of treatments for personality disorder as were included in this meta-analysis, so a further systematic review including only randomized trials is due.

Psychotherapy for borderline personality disorder (BPD)

Several psychosocial treatments for BPD have emerged over the past decade, with one US guideline recommending psychotherapy as the primary treatment for this condition.[7] It is impossible to recommend one specific therapy, because information from research remains inadequate. It has become clear not only that several treatments may be of use, but also that any one treatment by itself, is helpful in only about a half of all cases.[8] Also, there is general consensus that some of the nonspecific elements of psychotherapy may be as important in determining the success of a treatment as the specific techniques. We reviewed treatments shown to be moderately effective and concluded that they share certain common features.[9] They tend to (a) be well-structured, (b) devote considerable effort to enhancing collaboration, (c) have a clear focus, (d) be theoretically highly coherent to both therapist and patient, (e) be relatively long-term, (f) encourage a powerful attachment relationship between therapist and patient, and (g) be well integrated with other services available to the patient. While some of these features may seem to pertain more to a successful research study rather than a successful therapy, the manner in which clinical treatment protocols are constructed and delivered is probably as important in the success of treatment as the intervention itself.

(a) Dynamic psychotherapy

Dynamic psychotherapy has long been recommended for BPD and has now been modified to target the characteristic features of the disorder. Almost all of the first studies examined inpatient treatment using prospective one group pre- to post-test designs.[10] These studies failed to rule out other plausible causes of change, such as passage of time or subsequent outpatient treatment. Stone's[11] report of up to 20 years follow-up of 550 inpatients, most of whom had received some sort of psychosocial intervention, indicated that 66 per cent of patients were functioning well. However, a naturalistic 5-year follow-up of individuals receiving inpatient treatment at the Cassel Hospital in London indicates the need for caution in ascribing benefits to inpatient treatment.[12] Whilst longer term follow-ups are to be applauded, they are hard to interpret, because other therapies are often given subsequent to the original treatment.

Several nonrandomized trials of dynamic psychotherapy have been undertaken by Stevenson, Meares, and colleagues. In an open trial for 48 patients receiving twice-weekly interpersonal-psychodynamic outpatient therapy for 12 months,[13] 30 per cent of the treatment group no longer met DSM-III-R criteria for BPD, while the waiting-list group changed little. Cost–benefit analysis found significant reduction in costs, largely attributable to reduced inpatient stays. A replication study[14] also found significant reduction in symptom severity with the same treatment.

A randomized study of 38 patients with BPD compared an 18-month programme of partial hospitalization using mentalization-based treatment (MBT) with standard psychiatric care.[15] Mentalization entails making sense of the actions of oneself and others on the basis of intentional mental states such as desires, feelings, and beliefs. Outcome measures included frequency of suicide attempts, acts of self-harm, number and duration of inpatient admissions, use of psychotropic medication, and self-report measures including depression, interpersonal function, and social adjustment. After 6 months, patients given MBT showed a statistically significant decrease on all measures in contrast to the control group which showed limited change or deterioration over the same period. This was sustained at the end of the 18 months of treatment and showed further improvement on follow-up after another 18 months. Long-term follow-up 5 years after the initial treatment suggested that the differences between the groups continued, but general social function remained impaired in both groups.[16] The treatment was cost-effective and has been manualized,[17] but as yet the active components remain unclear, and it has not been shown that positive outcomes are correlated with an improvement in mentalizing. (An outpatient version of MBT is currently being evaluated for borderline and antisocial PD in a further randomized controlled trial.) Although promising, this treatment needs further validation by research carried out independently of the originators. Favourable data has recently become available on the effectiveness of a similar programme established in the Netherlands.[18]

Transference-focused psychotherapy (TFP) has also shown good results. In a cohort study,[19] 23 female borderline patients were treated for 12 months. Compared with the year before treatment, the number of patients who made suicide attempts decreased significantly, as did the medical risk and severity of self-injurious behaviour. Also, compared with the previous year, there were significantly fewer hospitalizations and fewer days of psychiatric hospitalization. However, one in five patients dropped out of treatment. A subsequent trial compared TFP, DBT, and supportive therapy.[20] Ninety patients (all but nine of whom were women) were randomized. At the end of 1 year of treatment, the groups did not differ on global assessment of functioning, social adjustment, scores of depression and anxiety, and measures of self-harm. TFP patients improved significantly more than those receiving DBT or supportive therapy on irritability and verbal and direct assault. Patients who received TFP improved most in reflective function, an operationalization of the mentalization construct,[21] but it is not known whether improved reflective function relates to treatment gains at follow-up.

Schema-focused therapy (SFT) has been compared with TFP.[22] Treatment was given by therapists with approved training in the treatment methods administered treatment to 88 patients with borderline personality disorder. In an 'intent to treat' analysis, patients who received TFP showed significantly less improvement than those who received schema-focused CBT over 3 years, and TFP was more expensive. Both groups showed improvement, but changes in the combined measure of outcome in the schema-focused therapy group were greater and more prolonged than in the TFP group. There are several reasons for caution regarding these conclusions. (i) Differences in outcome between the groups can be accounted for almost entirely by the larger dropout early in treatment of patients treated with TFP and disappear when 'completers' are compared. It would be valuable to know why more patients dropped out from TFP at an early stage than from SFT. (ii) Follow-up is required to determine whether treatment gains and group differences are maintained. (iii) In the duration of the treatment period, around 40 per cent of patients could be expected to improve without the treatment.[23] This study also raises the question of how successfully a treatment (TFP) from the US can be transported to a European context.

(b) Group psychotherapy

Noncontrolled studies of day hospital stabilization followed by outpatient dynamic group therapy indicate its utility for BPD.[24] Marziali and Monroe-Blum used group therapy focused on relationship management and without the milieu and social components. A randomized controlled trial found equivalent results between group and individual therapy, suggesting that group therapy is more cost-effective.[25] Further studies are needed to confirm their findings, especially since dropout rates were high.

(c) Cognitive analytical therapy (CAT)

CAT has been manualized for treatment of BPD, and many are enthusiastic about its effectiveness. There are some indications that it may be effective. In a series of 27 patients with borderline personality disorder treated with 24 sessions of CAT, half no longer met diagnostic criteria for personality disorder at 6-month follow-up.[26] More definitive statements about efficacy await results of a randomized trial in progress. Ryle (personal communication) has indicated that patients treated with CAT showed significant improvement on a range of clinical measures but reported no difference between people receiving CAT and those undergoing other psychological treatments. Thus, the effects may be nonspecific. A second randomized trial is in progress comparing CAT with 'best available standard care' for adolescent patients with borderline personality disorder.[27]

(d) Cognitive therapy

Cognitive behavioural formulations of BPD are diverse. In a model derived from 'standard' CBT and modified for personality disorders, Beck and associates[28] define personality in terms of patterns of social, motivational, and cognitive-affective processes, thereby moving away from a primary emphasis on cognitions. However, personality is considered to be determined by 'idiosyncratic structures', known as schemas, whose cognitive content gives meaning to the person, and these schemas are the cornerstone of cognitive formulations of BPD. Young[29] has developed a treatment programme for BPD based on early maladaptive schemas (EMS). These are stable, enduring patterns of thinking and perception that begin early in life and are continually elaborated. EMS are unconditional beliefs linked together to form the core of an individual's self-image. Challenge to these beliefs threatens the core identity which is defended with alacrity, guile, and desperation since activation of the schemas may evoke aversive emotions. The EMS give rise to 'schema coping behaviour', which offers the best adaptation to living that the borderline has found. Schema-focused therapy (SFT) is only just being evaluated, but its adherence to the general requirements of an effective treatment enumerated above suggests that it should be reasonably successful. The recent report comparing SFT with TFP[22] is described above, but SFT has yet to be shown to be more effective than treatment as usual. It is possible that TFP simply induced more negative effects in patients than SFT.

A small ($N = 34$), randomized controlled trial assessed brief cognitive therapy, linked to a manual and incorporating elements of dialectical behaviour therapy, in the treatment of recurrent self-harm in people with cluster B personality difficulties and disorders.[30]

Manual-assisted cognitive treatment (MACT) is a complex six-session treatment based on the theory that deliberate self-harm and suicide attempts stem from distorted cognitive schemas and coping skills deficits.[31] It incorporates elements of bibliotherapy, CBT, and DBT, as well as psychoeducation in relation to self-harm and suicide attempts, and a functional analysis of specific episodes. The treatment also involves strategies to regulate emotion, such as distraction, crisis planning, and problem-solving strategies. Cognitive restructuring strategies and management of negative thinking are incorporated in the second phase of the programme, which includes components for the management of substance abuse and relapse prevention. Its brevity makes MACT a potentially valuable intervention from a public health standpoint. In a clinical trial, 34 self-harm repeaters with a parasuicide attempt in the preceding 12 months were randomly allocated to MACT or treatment as usual (TAU). The rate of suicide acts was lower with MACT, and self-rated depressive symptoms also improved. The mean treatment time was 2.7 sessions, and the average cost of care was 46 per cent less with MACT. A subsequent larger study ($N = 480$) did not find evidence that time to repeat parasuicide was extended following MACT, although there was a decrease in the cost of care.[32] A randomized controlled trial with 104 people with borderline personality disorder with a longer period of treatment (up to 30 sessions) found significant benefit with regard to suicidal behaviour but a nonsignificant increase in emergency presentations in the cognitive behaviour therapy group.[33] The CBT arm used less resources, although no significant cost-effectiveness advantage was demonstrated. Those who received CBT showed less evidence of dysfunctional beliefs, lower state anxiety scores, and less positive symptom distress.

Systems training for emotional predictability and problem solving (STEPPS)[34] is a group treatment offered as an adjunct to other treatments rather than as a sole intervention. It is a 20-week manualized programme of psychoeducation and behavioural management focusing on maladaptive schemas and including both professional and family carers. Subjects are encouraged to continue their usual care, including individual psychotherapy, medication, and case management, and are required to designate a mental health professional who would provide ongoing care and could be reached in a crisis. Data from 52 patients suggests some reduction in impulsive and suicidal behaviour and some improvement on measures of depression, but no follow-up data is yet available. An RCT is currently ongoing. In Holland, a retrospective assessment of the experience of 85 patients enrolled in STEPPS groups[35] reported significant improvement on all subscales of the SCL-90, particularly those assessing anxiety, depression, and interpersonal sensitivity. Patients and therapists reported moderate to high levels of acceptance for the treatment in both studies. As most of the effective programmes for BPD are long-term and expensive, a short-term efficacious treatment will be of great value. However, even if found to be effective in a clinical trial, its effectiveness will depend on the nature of the treatment as usual.

(e) Dialectical behaviour therapy (DBT)

DBT is a special adaptation of CBT, originally used for the treatment of a group of repeatedly parasuicidal female patients with borderline personality disorder. DBT is a manualized therapy[36] which includes techniques at the level of behaviour (functional analysis), cognitions (e.g. skills training), and support (empathy, teaching management of trauma), with a judicious mix of ideas derived from Zen Buddhism. The initial aim of DBT is to control self-harm, but its main aim is to promote change in the emotional

dysregulation that is judged to be at the core of the disorder. Thus, the goal of DBT goes far beyond self-harm reduction. The first trial[37] involved 44 females with borderline personality disorder who had made at least two suicide attempts in the previous 5 years, with one in the preceding 8 weeks. Half were assigned to DBT and half to the control condition. Assessments were made during and at the end of therapy and at a 1-year follow-up. Control patients were significantly more likely to attempt suicide spent significantly more time as inpatients over the year of treatment (mean 38.8 and 8.5 days, respectively), and were significantly more likely to drop out of the therapies to which they were assigned (attrition 50 per cent versus 16.7 per cent, respectively).

DBT was less superior at the 1-year follow-up. Follow-up was naturalistic, because the morbidity of the group was thought to preclude termination of therapy at the end of the experimental period. At 6-month follow-up, DBT patients continued to show less parasuicidal behaviour than controls, though there were no between-group differences in days in hospital. At 1 year there were no between-group differences in suicidal behaviour, but DBT patients had had fewer days in hospital. Treatment with DBT for 1 year compared with treatment as usual led to a reduction in the number and severity of suicide attempts and decreased the frequency and length of inpatient admission. However, there were no between-group differences on measures of depression, hopelessness or reasons for living.

The widespread adoption of DBT for BPD and other PDs is a tribute to both the effectiveness of the treatment and its acceptability to patients and families. Several studies have replicated the original Linehan study. Turner observed a decrease in parasuicidal acts and deliberate self-harm at 6 and 12 months in 12 patients treated with DBT compared to 12 treated with client-centred therapy.[38] Koons and colleagues[39] compared DBT with outpatient treatment as usual in 28 participants and found decreases in frequency of parasuicidal acts and self-injury at 3 and 6 months. Bohus and colleagues[40] explored an inpatient adaptation of DBT and found significant improvement in deliberate self-harm but a higher dropout rate. Other studies, however, have shown DBT to be no better than other active treatments such as the 12-step programme for opioid dependence[41] or treatment as usual in a UK context.[42]

Because of severity, symptomatology, and high rates of co-occurring disorders, BPD affects family members, and interventions addressing the needs of family members have been developed using a DBT frame.[43] In a study of a 12-week community-based BPD family education programme, Family Connections (FC),[44] family members showed significant improvements on burden, grief and empowerment, and a reduction in depression. However, the long-term effect of family interventions on the patient's well-being remains to be demonstrated.

It is uncertain which are the active elements of DBT-individual psychotherapy whether skills training, phone consultation, or the consultation team. Two studies examining the process of change in DBT[41,45] had inconclusive results. Nevertheless, adding a DBT skills training group to outpatient individual (non-DBT) psychotherapy does not seem to enhance treatment outcomes. Given that DBT is described as primarily a skills training approach, this finding might indicate that the central skills training component of DBT may not be of primary importance. However, individual DBT focuses on the strengthening of skills learned in the skills groups, and trying to combine a skills training group with an individual therapy that ignores or pays minimal attention to skill strengthening, may invalidate what the patient has learned about utilizing learned skills in an attempt to cope with everyday functioning. Disagreement remains regarding the policy of not admitting patients to hospital, except for a minimum period, since some studies show that the time and structure of an inpatient setting can be used to apparently good effect.[46]

(f) Therapeutic community treatments

A therapeutic community (TC) may be defined as an intensive form of treatment in which the environmental setting provides the core means through which behaviour can be challenged and modified, essentially through group interaction and interpersonal understanding. Therapeutic communities are described in Chapter 6.3.9. Several studies have been completed at the Cassel Hospital, a tertiary referral centre with a psychosocial residential treatment programme which includes daily unit meetings, community meetings, structured activities, co-responsibility planning for the running of the therapeutic community and other structured activities, and formal psychoanalytic psychotherapy (individual and small group). Patients in two different specialist psychosocial programmes (step-down and long-term inpatient) and in general psychiatric treatment as usual (TAU) were assessed at 12 and 24 months. By 24 months, patients in the step-down condition showed significant improvements on all measures.[47] Patients in the long-term residential condition showed significant improvements in symptom severity, social adaptation, and global functioning but no changes in self-harm, attempted suicide, and readmission rates. Over the same period, patients in the TAU group showed no improvement on any variable except self-harm and hospital readmissions. All three groups were followed for 72 months after intake. The specialist step-down condition led to significantly greater change than either the solely inpatient model or TAU in most key dimensions of functioning, even 5 years after the 12 months hospitalization.[48]

While the study appears to show that a step-down programme leads to significant improvement, this conclusion should be qualified by the study's design limitations, including the lack of random assignment to the three conditions and the naturalistic geographical allocation. Overall, these findings are consistent with the general view that extended hospital admission, even to a psychotherapeutically oriented unit, may engender pathological dependency and regression. Against this, a prospective study of 216 patients with severe personality disorder treated at the Menninger Clinic in two psychoanalytically orientated inpatient units[49] found positive change at discharge and 1 year follow-up, with no evidence of deleterious effects due to regression and dependency. As there are now many treatments for personality disorder, therapeutic communities must be evaluated by comparison studies using acceptable experimental designs if they are to be considered a serious treatment approach.

(g) Nidotherapy

Nidotherapy is the name of a new form of systematic management for chronic and persisting mental disorders that concentrates on making environmental changes to bring about a better fit between person and environment.[50] All types of environmental change—physical, social, and personal—are considered relevant.

Nidotherapy is a collaborative treatment 'involving the systematic assessment and modification of the environment to minimize the impact of any form of the disorder on individual or on society'.[50] The word is derived from the Latin nidus or nest, an environment adapted to the object that is occupying it. The therapeutic aim is not to change the person but rather to change the environment so it better fits the person. The therapist should repeatedly question whether interventions given in the course of management incorporate analytic, cognitive, or behavioural psychotherapy—if they do, the intervention is no longer nidotherapy. Nidotherapy involves seeing the world from the patient's standpoint, joint planning of agreed environmental targets, a concentration on finding ways to improve social function, sharing responsibility for the programme so that the patient is the final owner, and the use of independent arbitrators to resolve disagreements about the change that is needed. The programme has several phases: identification of the boundaries of the nidotherapy, full environmental analysis, implementation of agreed environmental change, monitoring progress, and resetting targets. It addresses the patient's physical, social, and personal environments.

This therapy is at its early stages, with little other than anecdotal evidence.[51] Nevertheless, it represents a healthy contrast to the behavioural, cognitive, and analytic approaches which dominate most psychosocial interventions. There is no doubt that a person's environment is an important determinant of their experience and behaviour. This is accepted regardless of theoretical perspective, yet none of these perspectives offers a model of how environments should be modified to be more in line with the capacities of the person who exists within them. Modifying situational constraints may indeed be an independent and effective adjunct to other treatment protocols for personality disorder.

(h) Summary

The findings from the above studies suggest that some of the problems encountered by people with BPD may be amenable to talking/behavioural treatments. Several studies show that the effort by the recipient of care in sticking with the care package is rewarded by a decline in anxiety, depression, self-harm, hospital admission, and use of prescribed medication. A review using Cochrane methodology[52] concluded that the studies are too few and too small to justify full confidence in their results so that current therapies remain experimental. The reviewers suggest that people with BPD, if offered entry into a randomized study of therapies, may wish to consider that outcomes of both experimental and control groups will probably be better than those of standard care outside the trial.

Psychotherapy for avoidant personality disorder

Avoidant PD is the most prevalent personality disorder in the general population.[53] However, avoidant personalities tend to suffer quietly and reject emotional involvement with others. Nevertheless, their distress and functional impairment are probably comparable to those of people with BPD.

Studies of the efficacy of treatment for avoidant PD have been hampered by the close connection of the condition to social phobia. About half of avoidant patients also have a social phobia. There is controversy over the distinction between generalized social phobia and avoidant personality disorder, many authors regarding the latter as an exaggerated type of social phobia.[54] Many studies of social phobia include individuals with avoidant personality disorder. Few trials explicitly focus on this disorder. Most APD patients are treated using behavioural methods including exposure, social skills training, and systematic desensitization. Early studies suggested treatment gains that were maintained for up to 1 year after finishing active treatment.[55]

For phobic anxiety, there is some agreement that the cognitive behavioural strategies involved in traditional treatment tend sometimes to fail with the more severely disturbed patients because of resistance evoked by maladaptive personality traits.[56] Most studies report a pattern of outcome similar to that for generalized social phobia, with similar rate of gains to those without APD but lower end-point functioning,[57] though some report a slower rate of change.[58] Massion *et al.*[59] report a 5-year prospective study examining the impact of personality disorders in 514 patients in the Harvard/Brown anxiety research programme. The presence of a personality disorder reduced the probability of remission in social phobia by 39 per cent. Much of this reduction was explained by the presence of avoidant personality disorder. Alden[60] found that while there were significant improvements on a range of measures, only 9 per cent of patients treated for APD rated themselves as completely improved.

Therapeutic strategies for severe cases of APD should be tailored to the core pathology of the disorder—(1) a despised and unworthy sense of self connected to a punitive and blaming internal other; (2) an avoidant interpersonal style; (3) affect phobia; and (4) pseudomentalization (intellectualization and pretend relations). In achieving somewhat better functioning for avoidant patients, short-term dynamic psychotherapy and cognitive therapy have been shown to have good results.[61] For the more severely disturbed avoidant patients, day treatment programmes are probably more effective[62] than outpatient therapy, where they tend to ossify.[63] An RCT of cognitive behaviour therapy versus psychodynamic therapy for AVPD[64] found greater benefit from cognitive behaviour therapy. Interpersonal therapy and supportive-expressive therapy have also been tried. Patients are often reported to have made substantial gains in all treatment programmes,[65] but many patients who complete treatment still demonstrate a low level of social function.

Psychotherapy for antisocial personality disorder (ASPD)

There are few evaluated therapies for ASPD. The disorder is a common comorbid diagnosis for individuals with substance abuse problems,[66] and a small number of trials have contrasted those with ASPD alone with those with ASPD and depression in the treatment of substance misuse. Outcomes with antisocial personality disorder alone are often poorer than when ASPD is associated with depression. Three studies[67–69] have reported that ASPD alone was an obstacle in the treatment of substance misuse.

Several studies (though few controlled trials) have examined the impact of treatment on individuals within the penal system, at least some of whom, it is likely meet criteria for antisocial personality disorder. Most studies focus on specific categories of offending behaviour (e.g. sexual offences), or on prisoners with problem behaviours, particularly violence.[70] Some studies have focused on individuals who manifest callous-unemotional traits.[71]

There are a few observational studies of individuals detained in high security settings using group CBT,[72] individual and group CBT in the context of a therapeutic milieu,[73] or psychodynamic therapy.[74] While improvements in functioning are noted, there are severe methodological problems such as the absence of control groups and highly reactive measurement of outcome.

Studies reviewed by Warren and colleagues[70] suggest the usefulness of cognitive behavioural or (less frequently) systemic group-based interventions of various kinds for mentally disordered offenders. These include: seven studies of problem-solving skills training, five studies of anger/aggression management, and three studies aimed at social skills problems. Three studies of problem-solving training reported no statistically significant positive effects, while a further three reported some positive results with measures of outcome closely linked to the intervention. Anger management studies were predominantly negative with no statistically significant positive effects reported by the majority of studies and positive results reported only with quite reactive measures or within quasi-experimental studies. Social skills interventions also enjoyed mixed success with either no statistically significant positive effects being reported or positive effects observed only with reactive measures or in uncontrolled studies. While it is not possible to claim that psychotherapy is effective in addressing problems associated with ASPD, there is sufficient preliminary data to justify further inquiry and to counter the therapeutic nihilism that usually pervades this field.

Psychotherapy for other personality disorders

Individuals with **paranoid PD** are highly suspicious of other people, including doctors, and are often unable to acknowledge their own negative feelings towards others. This fact alone means that treatment is problematic, but given that other characteristics include concern that other people have hidden motives, expectation of being exploited by others, inability to collaborate, social isolation, emotional detachment, and overt hostility, treatment becomes nearly impossible. This constellation of characteristics means research on this group of patients is very difficult to conduct, since researchers are inevitably seen as having hidden motives. Whatever treatment method is employed, the first step is the development of a moderately trusting relationship. There is no research specifically investigating the outcome of psychotherapeutic treatment for paranoid personality disorder. However, a number of studies have included patients with paranoid PD in assessment of treatment. In general, the presence of paranoid PD diminishes the effectiveness of psychotherapeutic treatment for other co-occurring personality disorders.

Most of the interest in **schizotypal** personality disorder (STPD) has been in its relationship to schizophrenia, and hence there is little data on outcome of treatment using psychosocial methods. The Chestnut Lodge follow-up study yielded some data suggesting that patients who had schizophrenia plus STPD and had been treated within the psychotherapeutic milieu did slightly better than patients without STPD.[75] This unexpected finding has not been explained.

Winston *et al.*[76] treated 32 patients with a range of personality disorders, predominantly in DSM-III-R cluster C, using two forms of brief psychodynamic therapy (short-term dynamic psychotherapy or brief adaptational therapy) with 40 weekly sessions and waiting-list control. Because of ethical constraints, waiting-list subjects began treatment after 15 weeks. Although both therapies used psychodynamic techniques, brief adaptational therapy employed more cognitive strategies. Contrasted to waiting-list controls, treated patients showed moderate improvements in symptoms and target complaints at post-therapy. Though there were few between-treatment differences, there was more variance in outcome for the patients treated with short-term dynamic therapy, suggesting that for some patients the technique was unhelpful. Gains appeared to be maintained at 18-month follow-up. A larger trial of these therapies with mainly cluster C disorder[77] led to similar findings.

Conclusions

Effectiveness

It is difficult to reach definitive conclusions about the effectiveness of psychotherapy in the treatment of personality disorders, chiefly because RCTs of psychological treatments for PDs are relatively scarce. There is evidence that personality disorders impact negatively on treatments for Axis I presentations. However, the strength and specificity of this evidence varies both across Axis I diagnoses and also across personality disorders.

Historically, studies of psychotherapy for personality disorder have focused on cluster B PDs. Borderline personality disorder is the best studied of the PD diagnoses both from a biological and a psychotherapeutic perspective. The recent growth in the evidence base indicates the feasibility of coherent research programmes with a patient group usually seen as presenting substantial problems of engagement. Evidence from randomized trials suggests that structured treatments employing DBT or a psychodynamic approach are more effective than routine care. Data gathered so far support the use of both behavioural and psychodynamic interventions, while evidence for more short-term cognitive interventions is somewhat equivocal. Since most studies present data from open trials or case-series, only two approaches can be considered 'evidence-based'. The available randomized trials of DBT and psychodynamic treatment are methodologically sound but limited in the latter case by relatively small sample sizes; further, both sets of studies were conducted in the context of clear leadership from the developers. Because, for the most part, contrast was to routine care, it is difficult to ascertain whether outcomes are due to the structured nature of the programmes or their therapeutic orientation.

In relation to antisocial personality disorder, research into treatments has been centred on patients seeking help for substance abuse or individuals seen within the penal system for problems associated with ASPD. There is no clear evidence of treatment efficacy for any one form of intervention. Conclusions in relation to ASPD are limited not only by the small number of trials but also by the methodological limitations of these studies, particularly reactive measures of outcome and complex multifaceted interventions (such as therapeutic communities) where effective transportable components of the treatment package are impossible to disaggregate.

Avoidant personality disorder has received only a modest amount of research attention. Individuals presenting with APD would normally be offered social skills training or cognitive (exposure-based) techniques which are likely to be successful, but the generalization of improvements to other social contexts has not yet been well demonstrated.

Implications for clinical practice

With the exception of cluster B disorders, many individuals with PDs present to services for help with Axis I disorders. Clinicians should be aware that while treatments normally recommended for these problems can still work, they may do so to a lesser extent than would be expected if comorbid PD were not part of the clinical problem. However, they should not assume that this will definitely be the case. The situation is somewhat different in the case of the so-called dramatic PDs. Many of the features of the behaviour of individuals with borderline or antisocial PD create such distress for the people around them that clinicians tend to privilege these features over symptomatic states related to comorbid Axis I disorders when setting goals for treatment.

There is little doubt that psychological therapies have a place in the management of personality disorders, but the methodological shortcomings of the studies so far conducted limit our ability to make practical recommendations. Given the nature of the problem (supposedly enduring problems of character) the fact that few trials conduct follow-up over a longer period or include a contrast or control group over follow-up is a grave limitation of the database. Only two studies (Chiesa's 72 months follow-up and the MBT 5-year follow-up) have demonstrated that post-treatment differences are maintained for extended periods. Clinical experience, somewhat contradicted by recent follow-along data, suggests that many individuals may show periods of complete remission followed by episodes of great severity, and therefore it is crucial to investigate the impact of treatment on this often chronic and cyclic course. DBT has good evidence for its immediate impact on behavioural problems of impulsivity and suicidality, but the evidence for long-term benefit without further treatment is less clear. Psychodynamic approaches aim to change the way the person thinks, and there is some evidence that this approach has some impact on mood states and interpersonal functioning. However, this emphasis on cognitive change requires longer term therapy, with some indication of slower rates of change but perhaps also more sustained gains over follow-up. While shorter interventions are likely to have lower immediate costs, longer ones would be justified if they were more cost-effective in terms of service utilization and other cost-offsets. However, evidence for these kinds of effects is scarce.

Debate about particular forms of therapy may be less relevant than attention to nonspecific factors, especially in the treatment of individuals with cluster B PDs. We have seen that successful approaches usually emphasize the importance of structure, a coherent theoretical base, a higher intensity of treatment than for many Axis I disorders, and especially, attention to the unique set of problems presented by the individual, with a clear psychological formulation guiding treatment. Decisions about which treatment to opt for might be best made with regard to pragmatic considerations rather than argument about which particular course is 'best'. It may even be better to offer patients intensive treatment initially and to follow with short bursts of treatment over a long period of time. The competence and training of senior clinicians who can offer supervision will be especially important, as will the skill mix of staff and the resources (including staffing level) available to the service. These 'nonspecific' issues may be especially pertinent when considering the performance of evidence-based treatments in routine practice. Since systemic factors may be as relevant to success as the type of treatment offered, pragmatic trials would be useful to give an indication of the conditions required to implement evidence-based therapies in routine services.

Further information

Fonagy, P. (ed.) (2007) Special issue on treatment of personality disorder. *Journal of Mental Health*, **16**(1).

Gunderson, J.G. (2001) *Borderline personality disorder: a clinical guide*. American Psychiatric Publishing, Washington, DC.

Lenzenweger, M.F. and Clarkin, J.F. (2004). *Major theories of personality disorder* (2nd edn). Guilford, New York.

Roth, A. and Fonagy, P. (2005). Personality disorders. In *What works for whom? A critical review of psychotherapy research* (2nd edn) (eds. A. Roth and P. Fonagy), pp. 297–319. Guilford, New York.

References

1. Skodol, A.E., Gunderson, J., Shea, M.T., *et al.* (2006). The Collaborative Longitudinal Personality Disorders Study (CLPS): overview and implications. *Journal of Personality Disorders*, **19**(5), 487–504.
2. Zanarini, M.C., Frankenburg, F.R., Hennen, J., *et al.* (2006). The McLean Study of Adult Development (MSAD): overview and implications of the first six years of prospective follow-up. *Journal of Personality Disorders*, **19**(5), 505–23.
3. Grilo, C.M., Sanislow, C.A., Gunderson, J.G., *et al.* (2004). Two-year stability and change of schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders. *Journal of Consulting and Clinical Psychology*, **72**(5), 767–75.
4. Stone, M.H. (1990). *The fate of borderline patients: successful outcome and psychiatric practice*. Guilford Press, New York.
5. Perry, J.C., Banon, E., and Ianni, F. (1999). Effectiveness of psychotherapy for personality disorders. *The American Journal of Psychiatry*, **156**(9), 1312–21.
6. Leichsenring, F. and Leibing, E. (2003). The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders: a meta-analysis. *The American Journal of Psychiatry*, **160**(7), 1223–32.
7. Oldham, J.M., Phillips, K.A., and Gabbard, G.O. (2001). Practice guideline for the treatment of patients with borderline personality disorder. *The American Journal of Psychiatry*, **158**(Suppl. 10), 1–52.
8. Roth, A. and Fonagy, P. (2005). *What works for whom? A critical review of psychotherapy research* (2nd edn). Guilford Press, New York.
9. Bateman, A.W. and Fonagy, P. (2000). Effectiveness of psychotherapeutic treatment of personality disorder. *The British Journal of Psychiatry*, **177**, 138–43.
10. McGlashan, T. (1986). The chestnut lodge follow-up study III: long-term outcome of borderline personalities. *Archives of General Psychiatry*, **43**, 20–30.
11. Stone, M. (1993). Long-term outcome in personality disorders. *The British Journal of Psychiatry*, **162**, 299–313.
12. Rosser, R., Birch, S., Bond, H., *et al.* (1987). Five year follow-up of patients treated with in-patient psychotherapy at the Cassel Hospital for nervous diseases. *Journal of the Royal Society of Medicine*, **80**, 549–55.
13. Stevenson, J., Meares, R., and D'Angelo, R. (2005). Five-year outcome of outpatient psychotherapy with borderline patients. *Psychological Medicine*, **35**(1), 79–87.
14. Korner, A., Gerull, F., Meares, R., *et al.* (2006). Borderline personality disorder treated with the conversational model: a replication study. *Comprehensive Psychiatry*, **47**, 406–11.
15. Bateman, A.W. and Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: an 18-month follow-up. *The American Journal of Psychiatry*, **158**(1), 36–42.

16. Bateman, A.W. Fonagy P. 8-year follow-up of patients treated for borderline personality disorder—mentalization based treatment versus treatment as usual. *The American Journal of Psychiatry*. submitted.
17. Bateman, A.W. and Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: mentalization based treatment*. Oxford University Press, Oxford.
18. Bales, D., Andrea, H., Smits, M., *et al.* (2007). Mentalization based treatment in the Netherlands: first results. *Biannual meeting of the international society personality disorder*. Den haag, Netherlands.
19. Clarkin, J.F., Foelsch, P.A., Levy, K.N., *et al.* (2001). The development of a psychodynamic treatment for patients with borderline personality disorder: a preliminary study of behavioral change. *Journal of Personality Disorders*, **15**, 487–95.
20. Clarkin, J., Levy, K.N., Lenzenweger, M.F., *et al.* (2007). Evaluating three treatments for borderline personality disorder: a multiwave study. *The American Journal of Psychiatry*, **164**, 922–8.
21. Levy, K.N., Meehan, K.B., Kelly, K.M., *et al.* (2006). Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. *Journal of Consulting and Clinical Psychology*, **74**(6), 1027–40.
22. Giesen-Bloo, J., van Dyck, R., Spinhoven, P., *et al.* (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, **63**(6), 649–58.
23. Zanarini, M.C., Frankenburg, F.R., Hennen, J., *et al.* (2003). The longitudinal course of borderline psychopathology: 6-year prospective follow-up of the phenomenology of borderline personality disorder. *The American Journal of Psychiatry*, **160**(2), 274–83.
24. Wilberg, T., Friis, S., Karterud, S., *et al.* (1998). Outpatient group psychotherapy: a valuable continuation treatment for patients with borderline personality disorder treated in a day hospital? A 3-year follow-up study. *Nordic Journal of Psychiatry*, **52**, 213–22.
25. Marziali, E. and Monroe-Blum, H. (1995). An interpersonal approach to group psychotherapy with borderline personality disorder. *Journal of Personality Disorders*, **9**, 179–89.
26. Ryle, A. and Golynkina, K. (2000). Effectiveness of time-limited cognitive analytic therapy of borderline personality disorder: factors associated with outcome. *British Journal of Medical Psychology*, **73**(Pt 2), 197–210.
27. Ryle, A. (2004). The contribution of cognitive analytic therapy to the treatment of borderline personality disorder. *Journal of Personality Disorders*, **18**(1), 3–35.
28. Alford, B. and Beck, A. (1997). *The integrative power of cognitive therapy*. Guilford, New York.
29. Young, J.E. (1990). *Cognitive therapy for personality disorders: a schema-focused approach*. Professional Resource Exchange, Sarasota, Florida.
30. Evans, K., Tyrer, P., Catalan, J., *et al.* (1999). Manual-assisted cognitive-behaviour therapy (MACT): a randomized controlled trial of a brief intervention with bibliotherapy in the treatment of recurrent deliberate self-harm. *Psychological Medicine*, **29**(1), 19–25.
31. Schmidt, U. and Davidson, K. (2004). *Life after self-harm*. Routledge, London.
32. Byford, S., Knapp, M., Greenshields, J., *et al.* (2003). Cost-effectiveness of brief cognitive behaviour therapy versus treatment as usual in recurrent deliberate self-harm: a decision-making approach. *Psychological Medicine*, **33**(6), 977–86.
33. Davidson, K., Tyrer, P., Gumley, A., *et al.* (2006). A randomized controlled trial of cognitive behavior therapy for borderline personality disorder: rationale for trial, method, and description of sample. *Journal of Personality Disorders*, **20**(5), 431–49.
34. Blum, N., Pfohl, B., John, D.S., *et al.* (2002). STEPPS: a cognitive-behavioral systems-based group treatment for outpatients with borderline personality disorder—a preliminary report. *Comprehensive Psychiatry*, **43**(4), 301–10.
35. Freije, H., Dietz, B., and Appelo, M. (2002). Behandling van de borderline persoonlijk heidsstoornis met de VERS: de Vaardigheidstraining emotionele regulatiestoornis. *Directive Therapies*, **4**, 367–78.
36. Linehan, M.M. (1993). The skills training manual for treating borderline personality disorder. Guilford Press, New York.
37. Linehan, M.M., Heard, H.L., and Armstrong, H.E. (1993). Naturalistic follow-up of a behavioral treatment for chronically parasuicidal borderline patients. *Archives of General Psychiatry*, **50**, 971–4.
38. Turner, R.M. (2000). Naturalistic evaluation of DBT oriented treatment for BPD. *Cognitive Behaviour Practice*, **7**, 413–9.
39. Koons, C.R., Robins, C.J., Tweed, J.L., *et al.* (2001). Efficacy of DBT in women veterans with BPD. *Behavior Therapy*, **32**, 371–90.
40. Bohus, M., Haaf, B., Simms, T., *et al.* (2004). Effectiveness of inpatient dialectical behavioral therapy for borderline personality disorder: a controlled trial. *Behaviour Research and Therapy*, **42**(5), 487–99.
41. Linehan, M.M., Dimeff, L.A., Reynolds, S.K., *et al.* (2002). Dialectical behavior therapy versus comprehensive validation therapy plus 12-step for the treatment of opioid dependent women meeting criteria for borderline personality disorder. *Drug and Alcohol Dependence*, **67**(1), 13–26.
42. Feigenbaum, J.D., Fonagy, P., Pilling, S., *et al.* A pilot randomized control trial of the effectiveness of Dialectical Behavioural Therapy (DBT) for cluster B personality disorder. Submitted.
43. Hoffman, P.D. and Fruzzetti, A.E. (2007). Advances in interventions for families with a relative with a personality disorder diagnosis. *Current Psychiatry Reports*, **9**(1), 68–73.
44. Hoffman, P.D., Fruzzetti, A.E., and Buteau, E. (2007). Understanding and engaging families: an education, skills and support program for relatives impacted by borderline personality disorder. *Journal of Mental Health*, **16**(1), 69–82.
45. Shearin, E. and Linehan, M.M. (1992). Patient-therapist ratings and relationship to progress in dialectical behaviour therapy for borderline personality disorder. *Behavior Therapy*, **23**, 730–41.
46. Bohus, M., Haaf, B., Simms, T., *et al.* (2002). Effectiveness of inpatient dialectical behavioural therapy for borderline personality disorder: a controlled trial. *5th ISSPD European Congress on Personality Disorders—Abstracts*. Munich, Germany: 18.
47. Chiesa, M., Fonagy, P., Holmes, J., *et al.* (2004). Residential versus community treatment of personality disorders: a comparative study of three treatment programs. *The American Journal of Psychiatry*, **161**(8), 1463–70.
48. Chiesa, M., Fonagy, P., and Holmes, J. (2006). Six-year follow-up of three treatment programs to personality disorder. *Journal of Personality Disorders*, **20**(5), 493–509.
49. Gabbard, G.O., Coyne, L., Allen, J.G., *et al.* (2000). Evaluation of intensive inpatient treatment of patients with severe personality disorders. *Psychiatric Services*, **51**(7), 893–8.
50. Tyrer, P., Sensky, T., and Mitchard, S. (2003). Principles of nidotherapy in the treatment of persistent mental and personality disorders. *Psychotherapy and Psychosomatics*, **72**(6), 350–6.
51. Tyrer, P. and Kramo, K. (2007). Nidotherapy in practice. *Journal of Mental Health*, **16**(1), 117–29.
52. Binks, C.A., Fenton, M., McCarthy, L., *et al.* (2006). Pharmacological interventions for people with borderline personality disorder. *Cochrane Database of Systematic Review*, 2006(1), CD005653.
53. Samuels, J., Eaton, W.W., Bienvenu, O.J. III, *et al.* (2002). Prevalence and correlates of personality disorders in a community sample. *The British Journal of Psychiatry*, **180**, 536–42.
54. Rettew, D.C. (2000). Avoidant personality disorder, generalized social phobia, and shyness: putting the personality back into personality disorders. *Harvard Review of Psychiatry*, **8**(6), 283–97.
55. Renneberg, B., Goldstein, A.M., Phillips, D., *et al.* (1990). Intensive behavioral group treatment of avoidant personality disorder. *Behavior Therapy*, **21**, 363–77.

56. Alden, L.E., Taylor, C.T., Laposa, J.M., *et al.* (2006). Impact of social developmental experiences on cognitive-behavioral therapy for generalized social phobia. *Journal of Cognitive Psychotherapy: An International Quarterly*, **20**, 7–16.
57. Scholing, A. and Emmelkamp, P.M. (1999). Prediction of treatment outcome in social phobia: a cross-validation. *Behaviour Research and Therapy*, **37**(7), 659–70.
58. Oosterbaan, D.B., van Balkom, A.J., Spinhoven, P., *et al.* (2002). The influence on treatment gain of comorbid avoidant personality disorder in patients with social phobia. *The Journal of Nervous and Mental Disease*, **190**(1), 41–3.
59. Massion, A.O., Dyck, I.R., Shea, M.T., *et al.* (2002). Personality disorders and time to remission in generalized anxiety disorder, social phobia, and panic disorder. *Archives of General Psychiatry*, **59**(5), 434–40.
60. Alden, L.E. (1989). Short term structured treatment for avoidant personality disorder. *Journal of Consulting and Clinical Psychology*, **56**, 756–64.
61. Svartberg, M., Stiles, T.C., and Seltzer, M.H. (2004). Randomized, controlled trial of the effectiveness of short-term dynamic psychotherapy and cognitive therapy for cluster C personality disorders. *The American Journal of Psychiatry*, **161**, 810–7.
62. Karterud, S., Pedersen, G., Bjordal, E., *et al.* (2003). Day treatment of patients with personality disorders: experiences from a Norwegian treatment research network. *Journal of Personality Disorders*, **17**, 243–62.
63. Wilberg, T., Karterud, S., Pedersen, G., *et al.* (2003). Outpatient group psychotherapy following day treatment of patients with personality disorders. *Journal of Personality Disorders*, **17**, 510–21.
64. Emmelkamp, P., Benner, A., Kuipers, A., *et al.* (2006). Comparison of brief dynamic and cognitive-behavioural therapies in avoidant personality disorder. *The British Journal of Psychiatry*, **189**, 60–4.
65. Barber, J.P., Morse, J.Q., Krakauer, I.D., *et al.* (1997). Change in obsessive-compulsive and avoidant personality disorders following time-limited supportive-expressive therapy. *Psychotherapy*, **34**(2), 133–43.
66. Brooner, R.K., King, V.L., Kidorf, M., *et al.* (1997). Psychiatric and substance use comorbidity among treatment seeking opiate users. *Archives of General Psychiatry*, **54**, 71–80.
67. Woody, G.E., McLellan, T., Luborsky, L., *et al.* (1985). Sociopathy and psychotherapy outcome. *Archives of General Psychiatry*, **179**, 188–93.
68. Alterman, A.I., Rutherford, M.J., Cacciola, J.S., *et al.* (1996). Response to methadone maintenance and counseling in antisocial patients with and without major depression. *The Journal of Nervous and Mental Disease*, **184**(11), 695–702.
69. King, V.L., Kidorf, M.S., Stoller, K.B., *et al.* (2001). Influence of antisocial personality subtypes on drug abuse treatment response. *The Journal of Nervous and Mental Disease*, **189**(9), 593–601.
70. Warren, F., Preedy-Fayers, K., McGauley, G., *et al.* (2003). Review of treatments for severe personality disorder home office online report 30/03 (http://www.homeoffice.gov.uk/rds/pdfs2/rdsolr3003.pdf).
71. Hare, R.D., Hart, S.D., and Harpur, T.J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *Journal of Abnormal Psychology*, **100**, 391–8.
72. Quayle, M. and Moore, E. (1998). Evaluating the impact of structured groupwork with men in a high security hospital. *Criminal Behaviour and Mental Health*, **8**, 77–92.
73. Hughes, G., Hogue, T., Hollin, C., *et al.* (1997). First-stage evaluation of a treatment programme for personality disordered offenders. *Journal of Forensic Psychiatry*, **8**, 515–27.
74. Reiss, D., Grubin, D., and Meux, C. (1996). Young psychopaths in special hospital: treatment and outcome. *The British Journal of Psychiatry*, **168**, 99–104.
75. McGlashan, T.H. (1986). Schizotypal personality disorder. Chestnut Lodge follow-up study: VI. Long-term follow-up perspectives. *Archives of General Psychiatry*, **43**(4), 329–34.
76. Winston, A., Pollack, J., McCullough, L., *et al.* (1991). Brief psychotherapy of personality disorders. *Journal of Nervous and Mental Disease*, **179**, 188–93.
77. Winston, A., Laikin, M., Pollack, J., *et al.* (1994). Short-term psychotherapy of personality disorders. *The American Journal of Psychiatry*, **151**, 190–4.

4.12.7 Management of personality disorder

Giles Newton-Howes and Kate Davidson

Introduction

The treatment of personality disorders is a complex but rapidly evolving subject. It is to some extent described elsewhere in Chapters 3.3, 4.12.6, 4.13.1, 5.2.9, 6.3.5, 6.3.9 8.5.6, 9.2.4, 11.3.2, 11.16, and 11.17, and so this chapter excludes a full discussion of psychodynamic interventions, therapeutic communities, interventions for older people and the management of both adolescent and adult offenders.`

Methodological difficulties in evaluating the efficacy of treatment

The requirements for establishing whether a treatment is effective for personality disorders are much more exacting than those for mental state disorders. These can usefully be described under four headings:

- duration of treatment
- comorbidity
- adherence to treatment
- outcome measures.

(a) Duration of treatment

For most mental state disorders it is relatively easy to choose the period over which efficacy has to be demonstrated. In conditions that develop suddenly (e.g. panic), treatment trials could be for a very short time indeed. For others, particularly when maintenance treatment is being evaluated, at least 6 months may be necessary to establish continued efficacy. In the case of personality disorders, it has been thought that efficacy of treatment could not be judged adequately without at least a 2 to 3 year treatment phase. Personality disorder was regarded as being unlikely to change in the short-term. However, these ideas are changing in the light of evidence from longer-term follow-up studies of patients with personality disorder that show that these conditions do change over time in a consistent and predictable manner with substantial numbers of patients achieving full remission in the longer term.[1] If these longer-term follow-up studies are replicated, it would suggest that therapy should aim to accelerate the process of recovery. Determining what constitutes an adequate amount of therapy and over what length of time is an empirical question. More recent studies have offered treatment over 1 year with a 1 year follow-up

to examine maintenance of effect[2,3] but other studies have chosen lengthier treatment phases of up to 3 years[4] with some reporting continued therapy in the follow-up phase which does not allow the effect of maintenance of the original effect to be judged.[5] More recent studies, examining the effect of psychological treatments, have included a 1 year follow-up. The purpose of this follow-up period is to determine if treatment effects are maintained following the termination of treatment. Such a requirement is not a purist position; if a treatment for personality disorder appears to be effective over a shorter period, this may be due to change in a concurrent (comorbid) condition. In addition, if a treatment is to be judged efficacious in personality disorder its effects should be lasting beyond the active treatment component.

(b) Comorbidity

Comorbidity has been defined as 'the presence of any distinct clinical entity that has existed or that may occur during the clinical course of a patient who has the index disease under study'.[5] The key word here is 'distinct'. True comorbidity implies the presence of two completely separate disorders in the same person which are not causally related to each other in any way. Co-occurrence ranges from true comorbidity to the presence of the same disorder in two or more different forms.[6]

Comorbidity is the norm for most personality disorders both with other personality disorders or with mental state disorders. Borderline personality disorder is a major offender in this regard. Only about 1 in 20 of such disorders constitutes the pure condition,[7] and multiple comorbidity with four or more disorders is common.

In deciding on the efficacy of any treatment for personality disorder it is impossible to be certain whether observed improvement is in the personality disorder or in a comorbid condition. This problem is made worse because personality assessment is allegedly confounded, or 'contaminated', by the effect of a concurrent mental state disorder. Thus personality status apparently changes during the presence of a mental state disorder such as depression, only to return to the baseline normal subsequently.[8,9] Apparent improvement in a personality disorder following a treatment may be entirely due to improvement in a concurrent mental state disorder. However, this conclusion does not mean that *personality function*, as opposed to *personality disorder*, does not change. The underlying personality may remain stable, but if the setting and circumstances change, and this includes mental state changes, there can be marked changes in adjustment and so the manifestation of disorder will also change.[10]

In view of these problems, a treatment for a personality disorder should ideally be tested in those patients who have that personality disorder only. As these patients are uncommon and atypical, it is difficult to interpret the data from clinical trials.

(c) Adherence to treatment

People with personality disorders do not usually form good relationships with therapists. Although this is in keeping with their problems with relationships elsewhere, it can be a major problem in any form of therapy. The problem is particularly marked with psychotherapy, in which long-held views are challenged by the therapist. The consequence is that many patients dropout of care, and sometimes no amount of therapeutic skill can maintain them in care.[11] The failure to maintain prescribed treatment, in whatever form, is a constant handicap in accumulating an evidence base for interventions in personality disorder. Even within personality disorder there are differences between sub-groups. Most therapeutic trials have been inpatients with borderline personality disorder while those with schizoid, paranoid, histrionic, narcissistic, and antisocial personality disorders appear much less frequently. This is probably related to treatment attitudes. Borderline, anxious, and avoidant personality disorders contain a much higher proportion of treatment seeking (Type S) personality disorders as opposed to treatment resisting (Type R) ones, which are most prominent in those with antisocial, schizoid, and paranoid personalities.[12]

Any study of personality disorder is likely to have a large proportion of dropouts and this complicates the interpretation of the effects of treatment. The exception is when patients are treated in restricted settings such as prisons and other closed facilities,[13] but as these circumstances are abnormal it is difficult to generalize from them.

(d) Outcome measures

The choice of outcome measures is a problem in the assessment of all psychiatric disorders, but difficulty is particularly great in studies of personality disorders. These disorders affect both the individual and society, and a range of outcomes can be measured to cover these possibilities. Forensic psychiatrists and the general public usually consider that the outcome of mentally disordered offenders is best measured by the frequency of reoffending. This is an easily measured and reliable statistic, but it does not necessarily record symptomatic or personality change, and may be distorted by a range of other factors (e.g. patients who spend a long time in hospital or prison are not likely to reoffend). Changes in symptoms also have limited use since they may be a consequence of changes in mental state disorders quite independent of personality. Repeated measures of personality status also have disadvantages since, as noted earlier, they may be affected by changes with concurrent mental state disorder. Personality also changes with ageing irrespective of treatment.[14]

Because of these difficulties, global outcome measures are often used to determine the degree of improvement in personality disorders in long-term follow-up studies, although a battery of measurements is normally used in short-term treatment studies. Unfortunately, there is no standardized set of measures of global outcome. It is reasonable to take into account symptomatic change, social functioning, quality of life, incidents of societal conflict (e.g. police contacts), and reports from informants. Even these may not correctly reflect change in personality status. Thus a person whose personality disorder does not change in any basic way may find an environmental niche in which the personality disturbance does not manifest itself as conflict. Such a person would show improvement on all the items listed above, but the improvement would be a consequence of environmental change, not of personality alteration.

(e) Minimum requirements for establishment of efficacy

The evidence base for effectiveness of treatment in personality disorders is also exacting:

1 The treatment should be effective when used in the pure form of the personality disorder (in an explanatory trial) and subsequently in other forms of the disorder in which comorbidity is more common (pragmatic trial).

2 Efficacy cannot be established satisfactorily unless the treatment is tested in a randomized controlled trial.

3 A suitable control treatment or management needs to be tested against the experimental treatment.

4 Efficacy should only be determined after a period of at least 1 year because of the long duration of personality disorder. More recent randomized controlled trials in borderline personality disorder examining psychological therapies have met these more demanding criteria.[15,16]

Dynamic psychotherapy

This is discussed in chapter 4.12.6.

Cognitive therapies

(a) Cognitive analytical therapy

Cognitive analytical therapy combines cognitive and analytical approaches and has been applied to the treatment of personality disorders, particularly the borderline group.[17] The clinical manifestations of this condition are postulated to be a set of partially dissociated 'self-states' which account for the clinical features of this disorder. Such patients typically describe rapid switching from one state of mind to another, experiencing intense uncontrollable emotions or alternatively feeling muddled, or emotionally cut-off. Such 'dissociative states' (different from the conditions of similar name formerly linked to hysteria) are said to be activated by severe external threats and to be maintained by repetitions of threat and reinforcement by memories or situations which are similar to the original source of threat.

Cognitive analytical therapy is concerned with the identification of these different self-states and helping patients to identify 'reciprocal role procedures', or patterns of relationships which are learned in early childhood and are relatively resistant to change.[18] The patient is taught to observe and try to change damaging patterns of thinking and behaviour which relate to these self-states and to become more self-aware. The therapist's role is to gather information about the patient's experience of relationships and the different states he or she experiences, including interpretation where necessary. Although the standard measure of evidence of effectiveness, the randomized controlled trial, has still not yet been reported for this treatment it has gathered an impressive group of adherents and has become widely used and now has a good theoretical and pragmatic base.[18]

(b) Cognitive behaviour therapy

In its original form, cognitive therapy for depression was used to help the patient to identify and modify dysfunctional thoughts and beliefs through the use of specific cognitive techniques such as Socratic questioning. The focus of therapy was here and now the aim was to return the patient to his or her usual functioning by relieving current symptoms. The cognitive model of personality disorder does emphasize cognitive, emotional, and behavioural factors but the origins of personality problems are regarded as originating in the temperament of the child, childhood development, and experiences. Early infant attachment patterns, the child's internal working model of relationships, self-identity, self-worth, and the emotional availability of the infant's caregivers are central to how the child develops and these shape the adult self-identity, interpersonal relationships in adulthood, and behavioural and emotional coping responses.[19]

One of the first tasks of cognitive therapy in personality disorders is to gain an historical account of the patient's childhood development and background from which the therapist can derive a cognitive formulation linking past difficulties and presenting problems. Through the formulation, and understanding of the patient's view of self and others, unique core beliefs are identified that are linked to affect and to overdeveloped behavioural patterns that prevent the individual from functioning in an adaptive manner, particularly in interpersonal contexts. Therapy focuses on beliefs that concern core concepts about the self and others that have developed from childhood onwards and associated behaviours that have developed as coping strategies. The content and meaning of the beliefs have had an impact on past and present relationships and are likely to impact the therapeutic relationship. These beliefs, formed through negative, possibly abusive and neglectful experiences with others, are likely to have resulted in low self-esteem, hypersensitivity to criticism, and poor relationships with peers, caregivers and others in adolescence. Once a clear understanding of the content of patient's core beliefs and associated overdeveloped or compensatory behavioural patterns has been established, patients are encouraged to test out their beliefs and assumptions about others by learning new, more adaptive strategies for relating to others and to themselves. In borderline personality disorder, typically patients hold beliefs such as 'I am a bad and inadequate person' and 'others will abandon or reject me'. Having formed these beliefs through experiences in childhood, borderline patients, for example, may have learnt to avoid close relationships, are highly sensitized to signs of disapproval in others and have developed a punitive, self-critical style of thinking and behaviour, including self-harm. The emphasis in cognitive therapy is in developing new ways of thinking about self and others and in testing out new ways of behaving that are less self-defeating and more likely to improve the patient's interpersonal skills.[19] In comparison with the treatment of Axis I disorders, cognitive therapy with personality-disordered individuals takes more sessions and spans a longer time because the underlying problems are more pervasive and ingrained. There are other important elements of cognitive and related therapies, of which schema therapy and dialectical behaviour therapy are the most prominent. One of these differences is on the emphasis and attention paid to the therapeutic relationship. In cognitive therapy for personality disorder, more emphasis is placed on establishing and maintaining a therapeutic alliance, as interpersonal difficulties which occur in the patient's life outside therapy are also likely to arise within therapy. This is based on the hypothesis that the patient's core beliefs are consistent across a wide range of settings and therefore are also likely to be manifest in therapeutic relationships. Patients are therefore likely to be highly sensitive to signs of criticism and approval in their therapists. The models of treatment for personality disorder proposed by Beck and Freeman,[20] Davidson,[19] Young,[21] and Linehan[22] have in common an attempt to integrate biological and psychosocial factors. All models of treatment recognize the importance of building a secure therapeutic relationship and transference and countertransference issues in therapy are increasingly recognized as important mediators of the therapeutic process. These therapies utilize cognitive techniques to repair breakdown in communication that can occur during therapy.

Cognitive analytical therapy[18] also gives special attention to this aspect of therapy.

One of the goals of cognitive therapy with personality-disordered patients is to take advantage of these interpersonal difficulties in treatment by identifying and modifying the beliefs underlying them and, by extension, other relationships. Although people with personality disorders can recognize difficulties, they experience the problems as egosyntonic[23] (i.e. accepted as normal because they are an intrinsic part of usual functioning). As a result, alternative and potentially more adaptive beliefs about the self and others need to be identified and evaluated to see if they are indeed more adaptive and embraced as a consequence. These alternative more adaptive beliefs require to be systematically reviewed and reinforced, and new behaviours and ways of relating to others need to be practiced repeatedly if changes are to be consolidated. To achieve these changes, the therapist usually has to adopt a more directive approach than in cognitive therapy for depression and other Axis I disorders, and throughout will be more concerned with identifying and overcoming cognitive, emotional, and behavioural avoidance which maintains core beliefs.

(c) Other related psychological therapies

(i) STEPPS

Systems Training for Emotional Predictability and Problem Solving (STEPPS) is affiliated to the other cognitive psychotherapies. It was developed by Nancee Blum in Iowa and has been extended across several states within the United States and to the Netherlands. It has some of the elements of standard cognitive behaviour therapy and dialectical behaviour therapy and is a manualized programme involving 20 2 h weekly group meetings; with specific goals (or lessons) identified for each session.[24] A randomized controlled trial has just been completed and this shows significant gains in some areas compared with treatment as usual (Black, 2007, APA meeting, San Diego, USA).

(ii) Schema-focused therapy

Schema-focused therapy[21] is now becoming increasingly used in the treatment of borderline and antisocial personality disorders. It is a compendium of cognitive behaviour therapy, object relations theory, and gestalt therapy, and also involving what Young calls 'limited reparenting'. It is given in a relatively intensive form—two to three sessions a week for 1–2 years—but has been shown to be both more effective and cost-effective than transference-focused psychotherapy in a trial of treatments for borderline personality disorder.[28]

(iii) Dialectical behaviour therapy

The era of evidence-based therapy in personality disorder began with a formal trial of dialectical behaviour therapy, a form of cognitive behaviour therapy linked to skills training and detached acceptance (or mindfulness), was compared with treatment as usual in a group of repeatedly self-harming female patients with borderline personality disorder.[28,69] The hypothesis that dialectical behaviour therapy was effective in reducing self-harm was supported. Now several other randomized trials have taken place that show that DBT is particularly effective in reducing self-harm[4,29,30] though in another study, DBT improved hopelessness, depression, anger, and suicidal ideation but showed no difference in suicide attempts.[31]

This treatment has also been used systematically in the treatment of borderline personality disorder and those with comorbid substance abuse.[30] According to Linehan,[22] borderline personality disorder is primarily a dysfunction of emotional regulation which is assumed to have resulted from biological irregularities combined with certain dysfunctional environments. Others in contact with the patient are postulated as reinforcing this dysfunction by discounting or, in Linehan's preferred term, 'invalidating' their emotional experiences. Borderline patients are emotionally vulnerable and have difficulty in regulating patterns of responses associated with emotional states. The maladaptive behaviours which form part of the borderline syndrome can be viewed as either the product of emotional dysregulation or as attempts by the individual at regulating intense emotional states by maladaptive problem-solving strategies. Dialectical behaviour therapy, as its name suggests, contains within it the notion of opposites; common themes that emerge in therapy with borderline patients, such as acceptance of things as they are (so that there is no need for suicidal action), and change (from former maladaptive types of response) may appear incompatible but are synthesized in the therapy.

The essentials of dialectical behaviour therapy[22] are manualized weekly individual psychotherapy, group psychoeducational behavioural skills training, and telephone consultation when considered necessary. Therefore the content comprises a variety of problem-solving techniques including teaching the patient skills to help regulate emotions and tolerate distress, methods for validating the patient's perceptions, and behavioural and psychological versions of meditation skills. Therapists are also trained in case management. 'Core mindfulness skills' are also part of the treatment and involve teaching the patient to observe, describe, and take part in events and responses to events without dissociating from what is happening. The treatment encourages patients to take a non-judgemental approach to events and interactions and to do what is effective in situations rather than what they may feel is the 'right' thing to do.

Summary of cognitive therapies

The collective view of the effectiveness of cognitive and dialectical behaviour therapies is that they are undoubtedly effective Leichsenring and Leibing[32] examined 25 studies (but very few randomized controlled trials) and found cognitive behaviour therapy to be effective (effect size 1.0) but with psychodynamic therapies the effect size was larger (1.46), but the authors emphasized these results were preliminary and need updating. More randomized controlled trials have been carried out into treatments for personality disorder in the last 4 years than in the previous 50 years and further reviews are needed. To date we only have good evidence of effectiveness for borderline personality disorder, but cognitive behaviour therapy has also been shown to be effective in a randomized trial of avoidant personality disorder.[33]

Nidotherapy

Nidotherapy is 'the collaborative systematic assessment and modification of the environment to minimize the impact of any form of mental disorder on the individual or on society'.[25] It was developed specifically for the care of patients with multiple personality and major psychiatric pathology and is particularly focused on Type R (treatment resisting personality disorders). It makes no attempt to treat the patient directly but instead attempts to change the environment so there is a better fit between individual and

setting. It is normally carried out individually by a nidotherapist working independently from other clinical services.[26] It has been tested in a randomized trial in which the main gain was in cost savings as patients spent less time in hospital[27] and has also been used for antisocial personality disorder for the reduction of aggression and improvement in social functioning (http://www.controlled-trials.com (ISRCTN96256106)).

Therapeutic community treatments

There has been some confusion regarding the term 'therapeutic community'. It can apply to any form of environment created for a specific treatment but was really introduced and defined in its modern context by Maxwell Jones[34] who created a structure that ran completely counter to that of the traditional (authoritarian) mental hospital. This can be defined as a socially cohesive structure depending on intensive group treatments carried out by its residents. This approach, the democratic therapeutic community, differs from other forms of more coercive communities linked to criminal justice in the United States, often used for substance abusers, which are called concept therapeutic communities.[35,36] Therapeutic communities have very strong advocates and the complexity of the intervention makes it difficult to create a suitable comparison group to act as a control for a randomized trial; to date no such trial has been carried out.

Drug treatment

Drugs are often used for treating personality disorders although it is important to note that none are licensed for the treatment of these conditions. As with other forms of treatment, borderline personality disorder constitutes the largest group in which drug treatment is being used and is therefore worth examining separately. The evidence base for drug treatment is growing and there are now sufficient randomized controlled trials to evaluate and for other studies to be ignored in this review.

Borderline personality disorder

Borderline personality disorder is one of the most heterogeneous of all groups within the personality classification and includes extensive comorbidity with other personality disorders as well as with mental state disorders, particularly mood and stress-related disorders. This hinders the interpretation of Table 4.12.7.1, noting the problems of evaluation referred to at the beginning of this

Table 4.12.7.1 Summary of randomized controlled trials of drugs in the treatment of personality disorder (borderline except where indicated)

Drug group	Individual drug	Size of trial—*n* (ref)	Main outcome measures	Results
Tricyclic antidepressants@	Amitriptyline	61[37]	Depression, aggression, global improvement	Amitriptyline no better than placebo
Selective serotonin reuptake inhibitors (SSRIs)	Fluoxetine	40[38]	Aggression depression,	Fluoxetine superior to placebo
	Fluvoxamine (cross-over study)	38[39]	Mood shifts, aggression, impulsivity	Positive effects of fluvoxamine on mood only
Antipsychotic drugs (both first and second generation)	Haloperidol	61[37] and 108[40]	Depression, hostility, impulsiveness	Haloperidol effective in reducing depression and hostility in first study but not in second
	Olanzapine (wide-dosage range)	40[41]	Aggression, depression, anxiety	Significant improvement in an unrecognized scale (Clinical Global Improvement-BPD) only
	Olanzapine (mean 5.3 mg)	28[42]	'Random effects regression modelling of panel data'	Alleged improvement on composite measure (unsatisfactory)
	Aripiprazole 15 mg/day*	52[43]	Hostility, anger, depression, self-harm	Significant improvement in all measures and reduction in self-harm
Monoamine oxidase inhibitors (MAOIs)	Phenelzine (60 mg/day)	108[40]	Depression, hostility, impulsiveness, anxiety	Phenelzine superior to haloperidol and placebo for anger and hostility
Mood stabilizers and anti-convulsants	Carbamazepine (7 mg/day)†	20[44]	Depression, hostility	
	Sodium valproate (850 mg/day)*	30[45]	Depression, aggression, hostility	
	Sodium valproate*	91[46]	Aggression	
	Lamotrigine*	27[47]	Self-rated anger	
	Topiramate*	56[48]	Anger	

*Patients recruited by advertisement.

†Patients recruited as inpatients.

@A study involving mainly cluster C personality disorders compared the tricyclic antidepressant, dothiepin, with cognitive behaviour therapy and self-help over a 2-year period and showed significantly better response to dothiepin[49] (but study not included as after 2 years there were many deviations from the original protocol.

chapter. The studies are complex, yet the numbers are generally small, the period of treatment variable (8 weeks to 6 months), the dosage of drugs is usually flexible, and the outcomes manifold (and so variable that it is almost impossible to combine the data systematically). Positive publication bias and recruitment of volunteers by advertisement diminish the relationship between this patients group and those in clinical practice.

(i) Antidepressants

Tricyclic antidepressants do not appear to be effective in borderline personality disorder and, interestingly, do not help depressive symptoms preferentially, suggesting that there are subtle differences between the despair and emptiness of the borderline personality disorder and the anhedonia of depressive illnesses. Selective serotonin reuptake inhibitors (SSRIs), have been used in many trials (including some not cited here in which comparisons of inadequately small numbers have been made with psychological treatments) and provide some evidence of a positive effect on aggression and impulsiveness in antisocial personality disorder.[38] Monoamine oxidase inhibitors show reduction of anger and hostility in one trial but conventional risk management uggests that this treatment should only be used in exceptional circumstances.

(ii) Antipsychotic drugs

Although antipsychotic drugs have been tested in the treatment of borderline personality disorder more frequently than any other drug treatment, the results are equivocal. Haloperidol may be effective[37] in the short-term, however continuation studies do not show sustained improvement.[50] Olanzapine has been tested in three trials and, apart from showing consistent weight gain in all of these, has not shown any real evidence of benefit for any of the core symptoms of borderline personality disorder. The only trial showing substantial benefit was carried out with aripiprazole in symptomatic volunteers[43]; as aripiprazole has a complex pharmacological profile with 5-HT and dopamine agonist and antagonist actions this could represent a novel intervention and is worthy of replication.

(iii) Mood stabilizers

Borderline personality disorder is characterized by rapid swings in mood and emotional lability is a core feature. It is therefore not surprising that mood stabilizers have been used in its treatment. Again the results are equivocal and this subject urgently needs the benefit of a large independent trial. The most impressive result has been shown with the anticonvulsant drug, topiramate, which reduces aggression and hostility[48,51] but this finding needs replication. Lithium may also be effective in aggression but a satisfactory level of evidence is lacking.

Critique of drug treatment for borderline personality disorder

The evidence for the value of drug treatment has been influenced greatly by a guideline issued by the American Psychiatric Association in 2001.[49] This was a bold attempt to give clear recommendations to clinicians desperate to find a way through the fog of uncertainty with the abyss of suicide yawning on one side and iatrogenic poly-poly-pharmacy (the patients with this disorder are multiple consumers of prescribed drugs) on the other. There were four recommendations:

1 psychotherapy—both psychoanalytic/psychodynamic therapy, and dialectical behaviour therapy 'have been shown to have efficacy'
2 pharmacotherapy for 'affective dysregulation symptoms' should'be treated initially with a selective serotonin reuptake inhibitor or related antidepressant such as venlafaxine'
3 treatment of 'impulsive-behavioural dyscontrol symptoms' also suggests 'SSRIs are the initial treatment of choice'
4 'low-dose neuroleptics are the treatment of choice for "cognitive-perceptual" symptoms'.

These were criticized heavily at the time for going far beyond the evidence[52–54] and for some treatments, notably cognitive behaviour therapy, neglecting available evidence,[53] and further data accrued over the ensuing years has reinforced these concerns.[15,28,55] The recommendations concerning drug treatments are particularly suspect. There is no evidence worthy of the name that justifies the sub-grouping of borderline personality disorder into 'affective dyscontrol', 'impulsive-behavioural dyscontrol', and 'cognitive-perceptual' categories and these appear to be entities created to justify the use of particular drugs rather than provide a valid subdivision of a complex disorder. The creation of these essentially pseudo-diagnoses allows general conclusions to be made that all the drug treatments are efficacious and that 'taken together, the results of these studies suggest that the choice of medication can be guided as much by tolerability and safety as by symptom presentation'.[56] On the evidence analysed dispassionately this conclusion applies equally well to placebo, whose tolerability and safety are unparalleled.

Flamboyant group (cluster B, not borderline)

The evidence for drug use in other cluster B disorders is very limited. There may be a limited role for the use of mood stabilizers in reducing anger in dissocial personality. Lithium was found to be effective[57] and this action has been found in other settings,[58] but not reproduced in dissocial personality disorder. Carbamazepine has also been found to reduce impulsivity[59] but this result requires replication.

Odd eccentric group (cluster A)

No placebo-controlled, explanatory trials in this group have been conducted and no pragmatic trials provide evidence for drugs that remain in use. A mistrust of treatment given by others limits research and is probably an intrinsic part of the condition.

Anxious fearful group (cluster C)

The diagnostic overlap between cluster C personality disorders and neurotic disorders makes drawing conclusion about drug treatment for this group difficult. No clear explanatory trails have been conducted and the evidence of the effectiveness of antidepressants is confounded by the presence of neurotic disorder.[60,61]

Management

The management of borderline personality disorder can now be directed by a combination of the evidence and clinical judgement. It is clear from the randomized trials of both pharmacological and psychological treatment that an organized plan of care leads to improvement. This includes a consistent approach, a constancy of

personnel and adequate access to inpatient care during crisis situations. The cognitively based therapies are effective and pragmatic, providing a higher degree of input to ensure effectiveness. The intermittent addition of low-dose haloperidol or aripiprazole may assist with worsening impulsivity. A trial of topiramate for anger or SSRIs for impulsivity may also be of value. These need to be trialed for a specific period, 4 to 8 weeks, with a clear plan to review or stop if effectiveness is not clear or deteriorates over time. The risks of prescribing need to be carefully weighed against the potential benefits. Polypharmacy increases risk with no evidence of benefit. It is ethically important to ensure all patients have the capacity to consent to treatment, particularly for unlicensed drug use. Because the treatment of personality disorder is among the most complex of complex interventions[62] it is likely that several treatments, given in combination in a systematic way, are best suited to many of those who have more severe personality disorders.

The management of other personality disorders remains largely guided by clinical experience. The general principles of service organization remain the same, however decisions regarding psychotherapy and pharmacotherapy need to be tailored to individuals with an expectation of building a body of patient-based evidence, as other evidence is weak. Regular review of a management plan minimizes the likelihood of harm from any one course of action or the neglect of the patient's presentation.

Some management options can now be considered to have a negative risk: benefit ratio. These include behaviour therapy alone, tricyclic antidepressant therapy, and monoamine oxidase therapy.

Organization of services for personality disorder

It is likely that the organization of care for those with personality disorder has a much greater part to play than any specific single treatment in the clinical success of management but it is also fair to add that this conclusion is not based on excellent evidence. What is, however, clear from the randomized trials of both pharmacological and psychological treatment is that an organized plan of care leads to great improvement in those with personality disorder irrespective of the exact nature of the intervention.

Services for the management of personality disorder can include the 'sole practitioner' model (a single therapist seeing the patient), the 'divided function' model (in which part of the patient's care is taken over by specially trained staff whilst others look after other parts), and the 'specialist team' model (in which a specific personality disorder service takes on all aspects of care). The general conclusion is that the 'divided function or specialist team model is probably best for reducing risk and improving outcomes'.[63] The general principles of management apply to all personality disorders although the consequences of ignoring them will be greater for the flamboyant cluster of personality disorders than for others.

(a) Consistency

One of the reasons why those with personality disorders create so many problems in therapy is that they are highly sensitive to perceived criticism and are therefore able to detect any inconsistency in their treatment. Sometimes this is a way of deflecting attention away from fundamental problems associated with the personality disorder, but they are nonetheless effective in creating a screen that prevents other issues from being addressed. Clearly the fewer the people involved in care, the less are the chances of creating inconsistency. Keeping the number of main workers to a small number, and maintaining good communication, is a sound goal.

(b) Constancy

It is helpful to avoid changes in staff as far as possible. This is of particular relevance in the treatment of borderline personality disorder in which changes in therapists often re-enact the problems of loss and despair that the patient experiences so commonly in relationships.

(c) Adequate inpatient support

Staff involved in hospital care often believe that people with personality disorders should be kept out of hospital. This belief is based on the observation that such people exploit the opportunities offered by admission and create circumstances whereby it is difficult to discharge them. Much of this is opinion and not founded on evidence. Patients with comorbid mental state and personality disorders actually have better outcomes if they have a hospital-oriented programme of care than treatment in the community, whereas the opposite is true in the absence of personality disorder.[64] However, this work was carried out before specific services for personality disorder were set-up in England, the first country in the world to create such services.[65] The initial pilot service has just been evaluated and the results are encouraging, but the problem remains that those who wish to seek treatment from these services comprise only a small proportion of the total who suffer.[72]

Problems with comorbidity

Perhaps the most important error in management is failure to recognize personality disorder when other psychiatric disorders are more prominent and appear to be the only presenting problem. This is becoming recognized in the development of treatment protocols. This problem is encountered widely in the mental health services among people presenting to emergency psychiatric clinics,[73] in services for the homeless mentally ill,[74] and among heavy users of psychiatric services[75] and those with multiple admissions.[68] In all these settings, personality disorder is often not recognized early enough. This is to some extent understandable as the assessment of personality disorder is difficult in, for example, a busy emergency clinic. Nevertheless, failure to achieve a predicted response is often due to an earlier failure to detect the presence of personality disorder.

Conclusions

Some years ago that most sceptical of academic psychiatrists, Michael Shepherd, in referring to the contents of a book entitled *Recent Advances in Psychiatry* commented that the content was more accurately defined as 'recent activities', as 'advances' was too generous a word. Whilst not going quite as far as this in regard to advances in the treatment of personality disorder it is fair to add that the promise of effective therapies across the spectrum of personality dysfunction remains a long way off and we must be very careful not to oversell the evidence. The complexity of personality disorders often requires complex intervention, however, until we are confident that single treatments are effective the arguments for evaluating them in combination have to be very strong on theoretical grounds to justify the cost.

We are still at the stage in which explanatory trials (ones demonstrating efficacy under optimal conditions) are at least as necessary as pragmatic ones (demonstrating benefit in conditions of ordinary practice). These need to be carried out with adequate numbers of patients (at least 50 in each treatment arm rather than an artificially derived sample size to justify a small number) and with good independent assessments carried out by research workers who are masked as much as possible from disclosure of treatment. These requirements are exacting but can be achieved.[67] We also need better pragmatic trials of patients seen in ordinary mental health practice whose treatment and characteristics are both representative and from whom it is possible to generalize findings with confidence. Currently there are very few studies that satisfy this requirement; one recent study combining an educational intervention with problem-solving is an exception.[68]

Despite the caution of these statements these are exciting times in the management of personality disorder. We are no longer listening to the once powerful lobby that claimed that patients with these conditions should not be treated by psychiatric services, or to the pessimists that still regard these conditions as untreatable. We are in the equivalent position as those in the early 1950s who suddenly became aware of the possibility that powerful treatments for severe mental illness were ready and waiting to be used. They do indeed appear to be within reach, but we must use them wisely.

Acknowledgement

The authors would like to thank Professor Peter Tyrer who co-authored the original chapter on which this is based.

Further information

Sampson, M., McCubbin, R., and Tyrer, P. (eds.) (2006). *Personality disorder and community mental health teams, a practitioners guide*. John Wiley & Sons, Ltd., West Sussex. ISBN: 978-0-470-01171-3.

National Institute for Health and Clinical Excellence (UK): http://guidance.nice.org.uk/page.aspx?o=357063&c=91523

References

1. Zanarini, M.C., Krankenburg, F.R., Hennen, J., *et al.* (2003). Longitudinal course of borderline psychopathology: 6 year prospective study. *The American Journal of Psychiatry*, **160**, 274–83.
2. Davidson, K. (2007). *Cognitive therapy for personality disorders: a guide for clinicians* (2nd edn). Routledge, Hove.
3. Verheul, R., van den Bosch, L.M.C., Koeter, W.J., *et al.* (2003). Dialectical behaviour therapy with women with borderline personality disorder. 12-month randomised clinical trial in the Netherlands. *The British Journal of Psychiatry*, **182**, 135–40.
4. Linehan, M.M., Comtois, K.A., Murray, A. M., *et al.* (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*, **63**, 757–66.
5. Feinstein, A. (1970). The pre-therapeutic classification of comorbidity in chronic disease. *Journal of Chronic Diseases*, **23**, 455–62.
6. Tyrer, P. (1996). Comorbidity or consanguinity. *The British Journal of Psychiatry*, **168**, 669–71.
7. Fyer, M.R., Frances, A.J., Sullivan, T., *et al.* (1988). Co-morbidity of borderline personality disorder. *Archives of General Psychiatry*, **45**, 348–52.
8. Coppen, A.L. and Metcalfe, H. (1965). The effect of a depressive illness on MMPI scores. *The British Journal of Psychiatry*, **111**, 236–9.
9. Stuart, S., Simons, A.D., Thase, M.E., *et al.* (1992). Are personality assessments valid in acute major depression? *Journal of Affective Disorders*, **24**, 281–9.
10. Tyrer, P., Coombs, N., Ibrahimi, F., *et al.* (2007). Critical developments in the assessment of personality disorder. *The British Journal of Psychiatry*, **190**(Suppl. 49), s51–9.
11. Adler, G. (1979). The myth of the alliance with borderline patients. *The American Journal of Psychiatry*, **136**, 642–5.
12. Tyrer, P., Mitchard, S., Methuen, C., *et al.* (2003). Treatment-rejecting and treatment-seeking personality disorders: Type R and Type S. *Journal of Personality Disorders*, **17**, 265–70.
13. Soloff, P.H., George, A., Nathan, R.S., *et al.* (1986). Progress in pharmacotherapy of personality disorders: a double blind study of amitriptyline, haloperidol and placebo. *Archives of General Psychiatry*, **43**, 691–7.
14. Seivewright, H., Tyrer, P., and Johnson, T. (2002). Changes in personality status in neurotic disorder. *Lancet*, **359**, 2253–4.
15. Davidson, K., Norrie, J., Tyrer, P., *et al.* (2006). The effectiveness of cognitive behaviour therapy for borderline personality disorder: results from the BOSCOT trial. *Journal of Personality Disorders*, **20**, 450–65.
16. Linehan, M.M., Comtois, K.A., Murray, A.M., *et al.* (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*, **63**, 757–66.
17. Ryle, A. (1997). The structure and development of borderline personality disorder: a proposed model. *The British Journal of Psychiatry*, **170**, 82–7.
18. Ryle, A. and Kerr, I.B. (2002). *Introducing cognitive analytic therapy: principles and practice*. John Wiley, Chichester.
19. Davidson, K. (2007). *Cognitive therapy for personality disorders: a guide for clinicians* (2nd edn). Routledge, London.
20. Beck, A.T. and Freeman, A. (1990). *Cognitive therapy of personality disorders*. Guilford Press, New York.
21. Young, J.E., Klosko, J.S., and Weishaar, M.E. (2003). *Schema therapy: a practitioner's guide*. The Guilford Press, New York.
22. Linehan, M.M. (1992). *Cognitive therapy for borderline personality disorder*. Guilford Press, New York.
23. Alexander, F. (1930). The neurotic character. *International Journal of Psycho-analysis*, **11**, 291–311.
24. Blum, N., Pfohl, B., St. John, D., *et al.* (2002). STEPPS: a cognitive behavioral systems-based group treatment for outpatients with borderline personality disorder-apreliminary report. *Comprehensive Psychiatry*, **43**, 301–10.
25. Tyrer, P., Sensky, T., and Mitchard, S. (2003). The principles of nidotherapy in the treatment of persistent mental and personality disorders. *Psychotherapy and Psychosomatics*, **72**, 350–6.
26. Tyrer, P. and Bajaj, P. (2005). Nidotherapy: making the environment do the therapeutic work. *Advances in Psychiatric Treatment*, **11**, 232–8.
27. Tyrer, P. and Tyrer, S. (2008). Other treatments for persistent disturbances of behaviour. In *Cambridge textbook of effective treatments in psychiatry* (ed. P. Tyrer and K.R. Silk). Cambridge University Press, Cambridge.
28. Giesen-Bloo, J., van Dyck, R., Spinhoven, P., *et al.* (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, **63**, 649–58.
29. Verheul, R., van den Bosch, L.M.C., Koeter, W.J., *et al.* (2003). Dialectical behaviour therapy with women with borderline personality disorder. 12-month randomised clinical trial in the Netherlands. *The British Journal of Psychiatry*, **182**, 135–40.
30. Linehan, M., Dimeff, L., Reynolds, S., *et al.* (2002). Dialectical behavior therapy versus comprehensive validation therapy plus 12-step for the treatment of opioid dependent women meeting criteria for borderline personality disorder. *Drug and Alcohol Dependence*, **67**, 13–26.

31. Koons, C.R., Robins, C.J., Tweed, J.L., *et al.* (2001). Efficacy of dialectical behavior therapy in women veterans with borderline personality disorder. *Behavior Therapy*, **32**, 371–90.
32. Leichsenring, F. and Leibing, E. (2003). The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders: a meta-analysis. *The American Journal of Psychiatry*, **160**, 1223–32.
33. Emmelkamp, P.M.G., Benner, A., Kuipers, A., *et al.* (2006). Comparison of brief dynamic and cognitive-behavioural therapies in avoidant personality disorder. *The British Journal of Psychiatry*, **189**, 60–4.
34. Jones, M. (1953). The therapeutic community: a new treatment method in psychiatry. Basic Books, New York.
35. Campling, P. (2001). Therapeutic communities. *Advances in Psychiatric Treatment*, **7**, 365–72.
36. Rutter, D. and Tyrer, P. (2003). The value of therapeutic communities in the treatment of personality disorder: a suitable place for treatment? *Journal of Psychiatric Practice*, **9**, 291–302.
37. Soloff, P.H., George, A., Nathan, R.S., *et al.* (1986). Progress in pharmacotherapy of personality disorders: a double blind study of amitriptyline, haloperidol and placebo. *Archives of General Psychiatry*, **43**, 691–7.
38. Coccaro, E.F. and Kavoussi, R.J. (1997). Fluoxetine and impulsive-aggressive behaviour in personality disordered subjects. *Archives of General Psychiatry*, **54**, 1081–8.
39. Rinne, T., van den Brink, W., Wouters, L., *et al.* (2002). SSRI treatment of borderline personality disorder: a randomized, placebo-controlled clinical trial for female patients with borderline personality disorder. *The American Journal of Psychiatry*, **159**, 2048–54.
40. Soloff, P.H., George, A., Nathan, S., *et al.* (1993). Efficacy of phenelzine and haloperidol in borderline personality disorder. *Archives of General Psychiatry*, **50**, 377–85.
41. Bogenschutz, M.P. and George Nurnberg, H. (2004). Olanzapine versus placebo in the treatment of borderline personality disorder. *The Journal of Clinical Psychiatry*, **65**, 104–9.
42. Zanarini, M.C. and Frankenburg, F.R. (2001). Olanzapine treatment of female borderline personality disorder patients: a double-blind, placebo-controlled pilot study. *The Journal of Clinical Psychiatry*, **62**, 849–54.
43. Nickel, M.K., Muehlbacher, M., Nickel, C., *et al.* (2006). Aripiprazole in the treatment of patients with borderline personality disorder: a double-blind, placebo-controlled study. *The American Journal of Psychiatry*, **163**, 833–8.
44. De La Fuente, J.M. and Lotstra, F. (1994). A trial of carbamazepine in borderline personality-disorder. *European Neuropsychopharmacology*, **4**, 479–86.
45. Frankenburg, F.R. and Zanarini, M.C. (2002). Divalproex sodium treatment of women with borderline personality disorder and bipolar II disorder: a double-blind placebo-controlled pilot study. *The Journal of Clinical Psychiatry*, **63**, 442–6.
46. Hollander, E., Tracy, K.A., Swann, A.C., *et al.* (2003). Divalproex in the treatment of impulsive aggression: efficacy in cluster B personality disorders. *Neuropsychopharmacology*, **28**, 1186–97.
47. Tritt, K., Nickel, C., Lahmann, C., *et al.* (2005). Lamotrigine treatment of aggression in female borderline-patients: a randomized, double-blind, placebo-controlled study. *Journal of Psychopharmacology*, **19**, 287–91.
48. Nickel, M.K., Nickel, C., Kaplan, P., *et al.* (2005). Treatment of aggression with topiramate in male borderline patients: a double-blind, placebo-controlled study. *Biological Psychiatry*, **57**, 498–9.
49. American Psychiatric Association. (2001). Practice guideline for the treatment of patients with borderline personality disorder. *The American Journal of Psychiatry*, **158**(Suppl.), 1–52.
50. Cornelius, J., Soloff, P., Perel, J., *et al.* (1993). Continuation pharmacotherapy of borderline personality disorder with haloperidol and phenelzine. *The American Journal of Psychiatry*, **150**, 1843–8.
51. Loew, T.H., Nickel, M.K., Muehlbacher, M., *et al.* (2006). Topiramate treatment for women with borderline personality disorder—a double-blind, placebo-controlled study. *Journal of Clinical Psychopharmacology*, **26**, 61–6.
52. McGlashan, T.H. (2002). The borderline personality disorder practice guidelines: the good, the bad, and the realistic. *Journal of Personality Disorders*, **16**, 119–21.
53. Sanderson, C., Swenson, C., and Bohus, M. (2002). A critique of the American Psychiatric Practice Guideline for the treatment of patients with borderline personality disorder. *Journal of Personality Disorders*, **16**, 122–9.
54. Tyrer, P. (2002). Practice guideline for the treatment of personality disorders: a bridge too far. *Journal of Personality Disorders*, **16**, 113–8.
55. Weinberg, I., Gunderson, J.G., Hennen, J., *et al.* (2006). Manual assisted cognitive treatment for deliberate self-harm in borderline personality disorder. *Journal of Personality Disorders*, **20**, 467–82.
56. Zanarini, M.C. (2004). Update on pharmacotherapy of borderline personality disorder. *Current Psychiatry Reports*, **6**, 66–70.
57. Sheard, M.H., Marini, J.L., Bridges, C.I., *et al.* (1976). The effect of lithium on impulsive aggressive behavior in man. *The American Journal of Psychiatry*, **133**, 1409–13.
58. Tyrer, S.P., Walsh, A., Edwards, D.E., *et al.* (1984). Factors associated with a good response to lithium in aggressive mentally handicapped subjects. *Progress in Neuropsychopharmacology & Biological Psychiatry*, **8**, 751–5.
59. Cowdry, R.W. and Gardner, D.L. (1988). Pharmacotherapy of borderline personality disorder: alprazolam, carbamazepine, trifluoperazine and tranylcypromine. *Archives of General Psychiatry*, **45**, 111–9.
60. Deltito, J.A. and Stam, M. (1989). Psychopharmacological treatment of avoidant personality disorder. *Comprehensive Psychiatry*, **30**, 498–504.
61. Ansseau, M., Troisfontaines, B., Papart, P., *et al.* (1991). Compulsive personality as predictor of response to serotonergic antidepressants. *British Medical Journal*, **303**, 760–1.
62. Campbell, M., Fitzpatrick, R., Haines, A., *et al.* (2000). A framework for the design and evaluation of complex interventions to improve health. *British Medical Journal*, **321**, 694–6.
63. Bateman, A.W. and Tyrer, P. (2004). Services for personality disorder: organization for inclusion. *Advances in Psychiatric Treatment*, **10**, 425–33.
64. Tyrer, P., Manley, C., Van Horn, E., *et al.* (2000). Personality abnormality in severe mental illness and its influence on outcome of intensive and standard case management: a randomised controlled trial. *European Psychiatry*, **15**(Suppl. 1), 7–10.
65. National Institute of Mental Health (England). (2003). *Personality disorder: no longer a diagnosis of exclusion*. Department of Health, London.
66. Geller, J.L., Fisher, W.H., McDermeit, M., *et al.* (1998). The effects of public managed care on patterns of intensive use of inpatient psychiatric services. *Psychiatric Services*, **49**, 327–32.
67. Davidson, K.M., Tyrer, P., Gumley, A., *et al.* (2006). Rationale, description, and sample characteristics of a randomised controlled trial of cognitive therapy for borderline personality disorder: the BOSCOT study. *Journal of Personality Disorders*, **20**, 431–49.
68. Huband, N., Duggan, C., McMurran, M., *et al.* Social problem-solving plus psychoeducation for adults with personality disorder: a pragmatic randomised clinical trial. *The British Journal of Psychiatry*, **190**, 307–13.

69. Linehan, M.M., Armstrong, H.E., Suarez, A., *et al.* (1991). Cognitive–behavioral treatment of chronically parasuicidal borderline patients. *Archives of General Psychiatry*, **48**, 1060–4.
70. Zanarini, M.C. and Frankenburg, F.R. (2003). Omega-3 fatty acid treatment of women with borderline personality disorder: a double-blind, placebo-controlled pilot study. *The American Journal of Psychiatry*, **160**, 167–9.
71. Ansseau, M., Troisfontaines, B., Papart, P., *et al.* (1991). Compulsive personality as predictor of response to serotonergic antidepressants. *British Medical Journal*, **303**, 760–1.
72. Crawford, M., Price, K., Weaver, T., *et al.* (2007). Learning the lessons: an evaluation of pilot community services for adults with personality disorder. Department of Health, London.
73. Breslow, R.E., Klinger, B.I., and Erickson, B.J. (1996). Acute intoxication and substance abuse among patients presenting to a psychiatric emergency service. *General Hospital Psychiatry*, **18**, 183–91.
74. North, C.S., Thompson, S.J., Pollio, D.E., *et al.* (1997). *A diagnostic comparison of homeless and nonhomeless patients in an urban mental health clinic. Social Psychiatry and Psychiatric Epidemiology*, **32**, 236–40.
75. Kent, S., Fogarty, M., and Yellowlees, P. (1995). Heavy utilization of inpatient and outpatient services in a public mental health service. *Psychiatric Services*, **46**, 1254–7.

4.13

Habit and impulse control disorder

Contents

4.13.1 Impulse control disorders

Susan L. McElroy and Paul E. Keck, Jr.

This chapter first defines impulse control disorders, and then summarizes available research on the clinical features, epidemiology, psychiatric comorbidity, family studies, psychobiology, and treatment response of the most common of these conditions (except for pathological gambling, which is reviewed in Chapter 4.13.2).

Definitions of impulse control disorders

Historically, impulse control disorders have been broadly defined as harmful behaviours performed in response to irresistible impulses.[1] In DSM-IV, an impulse control disorder is defined as the failure to resist an impulse, drive, or temptation to commit an act that is harmful to the individual or to others.[2] DSM-IV also stipulates that for most impulse control disorders, the individual feels an increasing sense of tension or arousal before committing the act and then experiences pleasure, gratification, or relief at the time of committing the act. In the text describing these disorders, DSM-IV states that after the act, there may or may not be genuine regret, self-reproach, or guilt. In ICD-10,[3] these conditions are classified as habit and impulse disorders and defined as repeated acts that have no clear rational motivation, cannot be controlled, and that generally harm the patient's own interests and those of other people. ICD-10 further states that the behaviour is associated with impulses to action.

In DSM-IV, impulse control disorders are listed in a residual category, 'Impulse control disorders not elsewhere classified', which includes intermittent explosive disorder (IED), kleptomania, pyromania, pathological gambling, trichotill16omania, and impulse control disorders not otherwise specified (NOS). Examples of impulse control disorders NOS are compulsive buying disorder, repetitive self-mutilation, pathological skin picking, and onychophagia (severe nail-biting).[1] In ICD-10, habit and impulse disorders are also listed as a residual category. Similar to DSM-IV, it includes pathological gambling, pathological fire-setting (pyromania), pathological stealing (kleptomania), trichotillomania, other habit and impulse disorders (which includes IED), and habit and impulse disorder, unspecified.

It should be noted that with mounting research, the impulse control disorders are increasingly viewed as complex conditions sharing, in addition to irresistible impulses to perform harmful behaviours, features of trait impulsivity, trait compulsivity, and mood dysregulation, as well as obsessive compulsive mood, and addictive disorders.[1]

Intermittent explosive disorder

Definition and clinical features

Intermittent explosive disorder (IED) is defined in DSM-IV as several discrete episodes of failure to resist aggressive impulses that result in serious assaultive acts or destruction of property (criterion A). Also, the degree of aggression expressed during an episode is grossly out of proportion to any precipitating psychosocial stressors (criterion B) and the explosive episodes are not better accounted for by another mental disorder or due to the direct physiological effects of a substance or a general medical condition (criterion C). Varying definitions of IED based on the DSM-IV criteria have been proposed and used.[4,5] One important set of research criteria for IED, for example, allows for less severe aggressive episodes, such as recurrent verbal outbursts against others without physical aggression, but requires that the aggressive episodes be recurrent and associated with distress or dysfunction.[4] Although ICD-10 lists IED under 'Other habit and impulse disorders', it does not provide specific criteria for its diagnosis.

Regarding phenomenology, persons with IED describe their aggressive episodes as explosive, uncontrollable, unpremeditated, and brief; often provoked by minor stimuli; and associated with various psychological and physical symptoms, especially changes in mood, awareness, and autonomic arousal.[4,6] The frequency of

episodes depends in part on how the disorder is defined. In the National Comorbidity Survey Replication (NCS-R), where the DSM-IV A criteria of 'several' episodes was operationalized to be three or more lifetime attacks, persons with IED had a mean of 43 lifetime attacks.

Many persons with IED describe problems with chronic or trait anger and frequent 'subthreshold' aggressive episodes in which they manage to resist enacting aggressive impulses or express them with less destructive behaviours (e.g. screaming rather than assault).[4,6] These subthreshold episodes are similar to the anger attacks (sudden episodes of intense anger with autonomic arousal) often described in patients with mood (bipolar and depressive) disorders.[7]

Of note, the relationship between IED specifically and impulsive aggression in general remains unclear and the two phrases are not necessarily synonymous. In particular, like other impulse control disorders, the aggression of IED usually involves elements of lack of control (e.g. compulsivity) and affect dysregulation (extreme anger, irritability and/or mood instability) as well as impulsivity. Thus, IED may be one form of impulsive aggression.

Epidemiology and course

Once thought to be rare, recent research has shown that IED is common in both clinical and general population samples. In the most rigorous study to date, the NCS-R, lifetime and 12-month prevalence estimates of DSM-IV IED in the general population were 7.3 per cent and 3.9 per cent, respectively.[5] The lifetime and 12-month prevalences of more narrowly defined IED (in which three episodes in the same year were required for diagnosis) were 5.4 per cent and 3.5 per cent, respectively. The disorder is also likely to be common in forensic populations but data are not available.

IED is probably more common in males than females. In the NCS-R, 9.3 per cent of men versus 5.6 per cent of women met lifetime DSM-IV criteria for the disorder. IED begins in childhood or adolescence; follows a chronic or episodic course; and is associated with distress, morbidity (e.g. injuries), and social and occupational impairment.[4-6] For example, in the NCS-R, IED had a mean age of onset of 14 years, was persistent over the life course (with averages of 6.2–11.8 years with attacks), and was associated with substantial role impairment.[5] However, the prevalence of the disorder was significantly lower among persons 60 years and older (2.1 per cent).

Associated psychopathology and comorbidity

IED often co-occurs with other psychiatric disorders.[4–6] In the NCS-R, 81.8 per cent of respondents with lifetime DSM-IV IED met criteria for at least one other lifetime DSM-IV disorder. Specifically, IED was significantly comorbid with all DSM-IV depressive, anxiety, and substance use disorders assessed after controlling for age, sex, and race. It was also significantly comorbid with oppositional defiant disorder, conduct disorder, and attention-deficit/hyperactivity disorder.

Importantly, the boundaries between IED and other conditions characterized by episodic and/or impulsive aggression have not been clearly delineated. Indeed, the comorbidity between IED and both bipolar disorders and Axis II disorders was not assessed in the NCS-R.[5] Comorbidity with Axis II disorders was not determined because the prevalence of these disorders was not evaluated. Comorbidity with bipolar disorders was not assessed because of how IED was defined; cases of IED with lifetime mania or hypomania were excluded from analysis (number not specified) so that the prevalence of IED was not overestimated by cases of bipolar disorder with anger attacks. Of note, although the relationships between IED and both bipolar and cluster B personality disorders remain unclear, clinical studies suggest that patients with IED have high rates of these conditions.[4,6]

Family studies

Family studies suggest that relatives of probands with IED have high rates of impulsive violent behaviour, substance abuse, and possibly mood and other impulse control disorders.[4,6,8,9] In a family history study of patients with temper outbursts meeting the first two DSM-III criteria for IED, non-adopted patients were significantly more likely than adopted patients to have a family history of temper outbursts.[9] Of 25 subjects with DSM-IV IED evaluated via the family history method, 8 (32 per cent) of subjects had a first-degree relative with probable IED, 20 (80 per cent) had at least one first-degree relative with a substance use disorder, 14 (56 per cent) a mood disorder, and 14 (56 per cent) an impulse control disorder.[6] A blinded, controlled family history study using broadly-defined IED criteria found a significantly increased morbid risk of the condition in relatives of affected probands (26 per cent) compared with relatives of control probands (8 per cent).[4]

Psychobiology

Persons with impulsive aggression have been consistently found to have abnormalities in serotonergic function.[4] Although most studies included subjects with impulsive aggression and personality disorders, a few included persons with IED or possible IED. Thus, in a study of 58 violent offenders and impulsive fire-setters, 33 (57 per cent) of whom had DSM-III IED, lower cerebrospinal fluid (CSF) concentrations of 5-hydroxyindoleacetic acid (5-HIAA) were found in the impulsive offenders and fire-setters than in the non-impulsive offenders and normal control subjects.[10]

In a functional magnetic resonance imaging (MRI) study of response to social threat, 10 subjects with IED showed exaggerated amygdala reactivity and diminished orbitofrontal cortex activation to faces expressing anger compared with controls.[11] The authors noted these findings were similar to other disorders characterized by impulsive aggression, including borderline personality and bipolar disorders, and that they supported a link between a dysfunctional frontal-limbic network and aggression.

Treatment response

Clinical experience suggests that IED may be less responsive to insight-oriented and more responsive to cognitive behavioural therapies, particularly those stressing anger management.[4,6,12] Medications reported effective in definite or probable IED, some in controlled trials, include antiepileptics (e.g. phenytoin, carbamazepine, oxcarbazepine), antidepressants (e.g. tricyclics, serotonin reuptake inhibitors), mood stabilizers (e.g. lithium, valproate), β-blockers, psychostimulants, and even antiandrogens. Mood stabilizer monotherapy and antidepressant augmentation of mood stabilizers have both been reported to successfully treat IED and/or anger attacks in patients with bipolar disorders.[6,7] Antidepressants have been reported to be effective in anger attacks associated with

major depression.[12] Finally, serotonin reuptake inhibitors, mood stabilizers, antiepileptics, and antipsychotics may be effective for impulsive-aggressive behaviour in personality-disordered patients.[13]

Kleptomania

Definition and clinical features

Kleptomania is defined in DSM-IV as follows:

- recurrent failure to resist impulses to steal objects that are not needed for personal use or for their monetary value (criterion A);
- increasing sense of tension immediately before committing the theft (criterion B);
- pleasure, gratification, or relief at the time of committing the theft (criterion C);
- the stealing is not committed to express anger or vengeance and is not in response to a delusion or a hallucination (criterion D);
- the stealing is not better accounted for by conduct disorder, a manic episode, or antisocial personality disorder (criterion E).

In ICD-10, kleptomania (or pathological stealing) is defined as the repeated failure to resist impulses to steal objects that are not acquired for personal use or monetary gain.

An increasing number of studies have systematically examined the phenomenology of groups of people with DSM-defined kleptomania.[14–17] In these studies, most subjects described irresistible impulses or urges to steal, tension with the impulses, and tension relief either during or shortly after the act of theft (as required by the DSM criteria). Many subjects described the impulses as senseless, intrusive, uncomfortable, and uncontrollable. Many tried to resist the impulses with varying degrees of success. Some reported pleasurable feelings during the act of theft, often described as 'a rush,' 'a high,' or 'a thrill.' Most patients reported instances of impulsive stealing, but some also described premeditated stealing, the aim of which was sometimes to relieve the impulses to steal. Many subjects reported that they had lied to conceal their stealing. Some subjects developed rules for their stealing behaviour—for instance, stealing only from work or from certain types of shops (e.g. drug stores but not department stores), or stealing certain items but not others (e.g. jewellery but not clothing). Many subjects considered their stealing to be wrong, and many, but not all, reported guilt or remorse after stealing. Subjects who had been arrested for shoplifting reported that it had varying effects on their symptoms—some stopped stealing completely, some stopped for a limited amount of time, while others reported that their stealing was unaffected. Some stated they continued to steal once incarcerated.

These studies have also found that some subjects with apparent kleptomania report varying degrees of amnesia surrounding the act of stealing.[17] Many of these subjects deny impulses, tension, or relief with their thefts. Other subjects who are not amnesiac for their stealing episodes may also deny experiencing impulses, tension, relief, and/or pleasure. For these subjects, stealing appears to have become automatic or habit-like.[17]

Epidemiology and course

Kleptomania is presumed to be rare but its prevalence is unknown. Available studies suggest that only a small portion of shoplifters (from none to 8 per cent) represent true cases of kleptomania.[14] However, it has been argued that these rates may be spuriously low because psychiatric evaluations may not have always been sufficiently thorough, operational diagnostic criteria were rarely used, and kleptomania may have been under-represented in the samples due to selection bias (i.e. people with repeated apprehensions were more likely to be legally rather than psychiatrically referred). Also, kleptomania may be relatively common in clinical populations; it was the second most common lifetime impulse control disorder in a group of adult psychiatric inpatients assessed with a structured interview, present in 9.3 per cent of the sample.[18]

Kleptomania is probably more common in women than in men.[14–17] Many cases begin in adolescence or early adulthood, and often follow an episodic or a chronic course.

Associated psychopathology and comorbidity

Clinical studies show that kleptomania often co-occurs with other Axis I psychiatric disorders, including mood, anxiety, substance use, eating, and impulse control disorders.[14–18] In the only controlled study,[16] 10 patients with kleptomania had significantly higher rates of comorbid psychiatric disorders, particularly mood disorders, other impulse control disorders, and substance abuse or dependence (mainly nicotine dependence), than 29 psychiatric comparison patients and 60 patients with alcohol abuse or dependence. Several studies, including the one controlled study, found especially high rates of bipolar disorders.[15–17] Conversely, high rates of kleptomania have been found in women with eating disorders[14] and patients with depressive disorders.[19]

Preliminary data suggest patients with kleptomania may also have high rates of certain Axis II disorders.[17] However, the relationship between kleptomania and antisocial personality disorder is not understood.

Family studies

Uncontrolled studies suggest kleptomania may be associated with increased familial rates of mood, substance use, anxiety, and possibly impulse control disorders.[17] For example, of 103 first-degree relatives of 20 patients with DSM-IIIR-defined kleptomania evaluated blindly by the family history method, 22 (21 per cent) had a major mood disorder, 21 (20 per cent) had a substance use disorder, and 13 (13 per cent) had an anxiety disorder, including seven (7 per cent) with obsessive compulsive disorder[15] Also, two (2 per cent) had apparent kleptomania. However, a controlled family study found similar rates of kleptomania in first-degree relatives of probands with obsessive compulsive disorder and those of control probands.[20]

Psychobiology

Preliminary research suggests kleptomania may be associated with serotonergic and frontal lobe dysfunction. In one study, the number of platelet serotonin transporters, evaluated via [3H] paroxetine binding, was lower in 20 patients with obsessive-compulsive related disorders, including five patients with kleptomania, than in 20 healthy control subjects.[17] In another study, 10 females with DSM-IV kleptomania were more likely than controls to have decreased white matter microstructural integrity in inferior frontal brain regions when evaluated with diffusion tensor imaging.[21]

Treatment response

Although no controlled psychological treatment studies of kleptomania have been published, various types of cognitive behavioural therapy may be effective.[12,17] There are also successful reports of the use of psychodynamic psychotherapies, but there are negative reports as well.[12,15]

Medical treatments with antidepressant, mood-stabilizing, or anxiolytic properties have been reported to be effective in kleptomania, primarily in case reports and case series. These treatments include tricyclics, serotonin reuptake inhibitors, trazodone, lithium, valproate, electroconvulsive therapy, and benzodiazepines.[12,15,17] There are also reports of patients with kleptomania responding to the opioid antagonist naltrexone and the antiglutamatergic agent topiramate.[12,17]

However, in the only controlled pharmacotherapy study of kleptomania published to date, an open-label trial of escitalopram treatment in 24 subjects followed by double-blind discontinuation in 15 of 19 responders, there was no difference in response rate (defined as greater than a 50 per cent decrease in theft episodes per week) between subjects receiving escitalopram (3 [43 per cent]) and those receiving placebo (4 [50 per cent]).[22]

Pyromania

Definition and clinical features

Pyromania is defined in DSM-IV as follows: deliberate and purposeful fire-setting on more than one occasion (criterion A) that is associated with tension or affective arousal before the act (criterion B), fascination with, interest in, curiosity about, or attraction to fire and its situational contexts (criterion C), and pleasure, gratification, or relief when setting fires, or when witnessing or participating in their aftermath (criterion D). Also, the fire-setting is not done for monetary gain, as an expression of sociopolitical ideology, to conceal criminal activity, to express anger or vengeance, to improve one's living circumstances, in response to a delusion or hallucination, or as a result of impaired judgement (criterion E), and is not better accounted for by conduct disorder, a manic episode, or antisocial personality disorder (criterion F). In ICD-10, pyromania (or pathological fire-setting) is defined as multiple acts of, or attempts at, setting fire to property or other objects, without apparent motive, and by a persistent preoccupation with subjects related to fire and burning. The essential features are as follows:

- repeated fire-setting without any obvious motive such as monetary gain, revenge, or political extremism
- an intense interest in watching fires burn
- reported feelings of increasing tension before the act, and intense excitement immediately after it has been carried out.

Although the authors were unable to locate any systematic reports of a group of people with pyromania by either of the above criteria sets, there are numerous case reports and case series of people with repetitive fire-setting behaviour who would probably meet these criteria for pyromania. For example, in what is still probably the largest study of pathological fire-setting, in 1951, Lewis and Yarnell[23] evaluated 1 145 of 2 000 American case records of males 16 years of age and older from the National Board of Fire Underwriters (selection criteria were otherwise not clearly specified). They concluded that 688 of these males were best classified as 'pyromaniacs' as 'they set fires for no practical reason and received no material profit from the act, their only motive being to obtain some sort of sensual satisfaction'. Lewis and Yarnell did not provide quantitative data summarizing these 688 cases, but stated that 50 of the subjects 'approached true pyromania', in that they were able to give a 'classical description of the irresistible impulse'. Specifically, before they set fires, these subjects described 'mounting tension; . . . restlessness; the urge for motion; . . . conversion symptoms such as headaches, palpitations, ringing in the ears, and the gradual merging of their identity into a state of unreality'.

Epidemiology and course

The prevalence of pyromania is unknown.[19] Although there are numerous studies of fire-setting behaviour, few of these studies systematically assessed pyromania in their subjects. Those that did use variable definitions of pyromania and reported widely discrepant rates. For example, in their 1951 study of 1 145 adult males with pathological fire-setting, Lewis and Yarnell[23] reported that 688 (60 per cent) could be classified as having broadly defined pyromania, but only 50 (4 per cent) as having the 'true' disorder.

Pyromania is probably more common in males than females and usually begins in adolescence or early adulthood.[19,23] How often childhood fire-setting represents pyromania is unknown. Clinical descriptions indicate that the course of pyromania may be episodic or chronic, but its course into old age is unknown.

Associated psychopathology and comorbidity

There are no studies of the psychiatric comorbidity of a group of people with well-defined pyromania. However, there are case reports of people with apparent pyromania who have comorbid mood, obsessive compulsive, eating, paraphilic, and possibly psychotic disorders.[23] In addition, impulsive fire-setters have been reported to have high rates of mood, substance use, and cluster B personality disorders, as well as suicide attempts.[10]

Family studies

There are no family history studies of pyromania, but studies of impulsive fire-setters suggest elevated familial rates of substance use disorders.[8]

Biological studies

There are no biological studies of pyromania, but studies of impulsive fire-setters suggest they have abnormalities in central serotonergic neurotransmission.[10]

Treatment response

There are no systematic treatment studies of pyromania.[12,19] There is one case of an 18-year-old male with DSM-IV pyromania and a chief complaint of 'feeling addicted to setting fires' who responded to the combination of topiramate and cognitive behaviour therapy.[24] There is also a case report of two men in both of whom pyromania appeared to be part of a paraphilia and responded to antiandrogen medication.[25]

Clinical reports on the treatment of pathological fire-setting in general stress use of various psychological interventions (e.g. cognitive behavioural, psychoeducational, supportive, and insight-oriented) to address the fire-setting behaviour, and appropriate

treatment of any comorbid psychiatric disorders. Preliminary data suggest combined psychosocial treatment approaches may be helpful in reducing further fire-setting behaviour, at least in juveniles.[19]

Trichotillomania

Definition and clinical features

Trichotillomania is defined in DSM-IV as follows:

- recurrent pulling out of one's hair resulting in noticeable hair loss
- an increasing sense of tension immediately before pulling out the hair or when attempting to resist the behaviour
- pleasure, gratification, or relief when pulling out the hair
- the disturbance is not better accounted for by another mental disorder and is not due to a general medical condition (e.g. a dermatological condition)
- the disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

In ICD-10, trichotillomania is defined as noticeable hair-loss due to a recurrent failure to resist impulses to pull out hairs. ICD-10 further states the hair-pulling is usually preceded by mounting tension and is followed by a sense of relief or gratification and that the diagnosis should not be made if there is a pre-existing inflammation of the skin, or if the hair pulling is in response to a delusion or a hallucination.

Hair is most often pulled from the scalp but also from the eyelashes, eyebrows, face, axilla, arms, legs, abdomen, and pubis.[26,27] Extracted hair may be chewed or swallowed. Medical complications include trichobezoars (hairballs that form in the stomach) and, uncommonly, obstruction or perforation of the stomach or bowel.

Some authorities have argued that both the DSM-IV and ICD-10 criteria for trichotillomania are too narrow, noting that patients with distressing hair pulling behaviour, especially children, may not always experience impulses and/or tension before hair pulling or relief with or after hair pulling.[26–28] Indeed, for some persons, the hair pulling may be automatic or habit-like and not associated with urges, tension, or relief. In addition, the hair loss may not be noticeable. For this reason and because hair pulling is a self-grooming behaviour, some authorities have argued that trichotillomania should be grouped with other self-grooming behaviours that may become problematic (e.g. skin picking and nail biting, which are discussed later) into a family of grooming disorders or body-focused impulse control disorders.[20,26,27]

Epidemiology and course

The prevalence of DSM-IV or ICD-10 defined trichotillomania is unknown, but survey studies suggest between 0.5 per cent—3.5 per cent of college students report problematic hair pulling.[26,27]

Clinical studies indicate trichotillomania is more common in females than in males; may begin in childhood, adolescence, or adulthood; and may have an episodic or chronic course.[26,27] Spontaneous remissions may occur, particularly in children with recent onset of the disorder.

Associated psychopathology and comorbidity

Trichotillomania often co-occurs with mood, anxiety, eating, substance use, and other impulse control disorders in clinical samples of adults.[26,27] It may also co-occur with various personality disorders, with histrionic, borderline, and obsessive compulsive commonly being cited.[26,27] In paediatric samples, trichotillomania is similarly associated with mood, anxiety, and substance use disorders.[26]

Family Studies

Preliminary family research suggests that trichotillomania may be associated with increased rates of obsessive-compulsive and grooming disorders among first-degree relatives.[26] A controlled family study of 22 probands with compulsive hair pulling, 17 (77 per cent) of whom met DSM-III-R criteria for trichotillomania, found that depression, alcoholism, drug abuse, obsessive compulsive disorder, and antisocial personality disorder were significantly more frequent among the first degree relatives of hair pullers than relatives of control probands.[28] Conversely, another controlled family study found significantly higher rates of grooming disorders (trichotillomania as well as pathological skin picking, pathological nail biting, and impulse control disorder NOS) in first-degree relatives of probands with obsessive compulsive disorder than in first-degree relatives of control probands.[20]

Psychobiology

Neuroimaging studies in subjects with trichotillomania have shown hyperactivity in the cerebellum and right superior parietal lobe as well as possible structural abnormalities of the left putamen, left inferior frontal gyrus, right cuneal cortex, and cerebellum.[26,27,29] Neuropsychological abnormalities found in trichotillomania patients have included increased error rates in spatial processing, divided attention, nonverbal memory, and executive functioning.[26,29]

Genetic studies suggest trichotillomania may be associated with sequence variants in the slit and trk-like 1 (SLITRK1) gene and the T10ZC polymorphism of the serotonin receptor 2A gene.[30,31]

Treatment response

There are a few published reports of the successful treatment of trichotillomania with insight-oriented psychotherapies, but there are many successful reports with various behavioural-based therapies.[12,26,27] Indeed, to date, four randomized, controlled studies supporting the effectiveness of behavioural treatments in adult trichotillomania have been published. In these studies, habit reversal was more effective than negative practice training (N=34); CBT was more effective than clomipramine and placebo (N=16); behaviour therapy was superior to fluoxetine and a wait-list control (N=43); and acceptance and commitment therapy plus habit reversal was superior to wait-list control (N=25).[12,26,27] In addition, one randomized, controlled trial found a cognitive behavioural therapy package of awareness training, stimulus control, and habit-reversal training was superior to minimal attention control in paediatric trichotillomania.[27]

Although many case reports and open trials describe successful treatment of trichotillomania with various serotonin reuptake inhibitors, controlled studies have yielded mixed results.[12,26,27] Two small (N=13 and N=12) double-blind, crossover trials found

clomipramine was superior to desipramine and equivalent to fluoxetine (which had beneficial effects), respectively, in reducing hair-pulling symptoms. In contrast, 2 slightly larger (N=21 and N=23) placebo-controlled, double-blind crossover studies of fluoxetine (both up to 80 mg/day) in adult chronic hair pullers found fluoxetine was not superior to placebo in suppressing hair-pulling symptoms.[12,26,27] In addition, as noted earlier, in the two controlled comparisons which found cognitive behaviour therapy or behaviour therapy superior to clomipramine and fluoxetine in 16 and 43 patients with trichotillomania, respectively, neither drug was superior to placebo in reducing trichotillomanic symptoms.[12]

Despite these negative studies, thymoleptics other than serotonin reuptake inhibitors have been reported to be effective for trichotillomania in case reports and case series. These include antidepressants such as imipramine, amitriptyline, isocarboxazid, trazodone, mianserin, and bupropion; the mood stabilizers lithium, olanzapine, and quetiapine; and various antimanic antipsychotics—used as either monotherapy or adjunctively with serotonin reuptake inhibitors.[12,26,27]

Other medications described as being effective for trichotillomania, primarily in case reports or case series, are buspirone, fenfluramine, topiramate, inositol, the progestin levonorgestrel, and naltrexone.[12,26,27,32,33] Naltrexone was reported superior to placebo in one small controlled trial which has only been presented in abstract form.[12] Case reports also describe the successful use of topical steroid ointments in combination with psychotropics when skin is irritated.[12]

Compulsive buying disorder

History and clinical description

Although compulsive buying disorder (also called compulsive shopping, buying mania, and oniomania) is not classified in DSM-IV or ICD-10 as a mental disorder, diagnostic criteria have been proposed.[34,35] These include being frequently preoccupied with buying or subject to irresistible, intrusive, and/or senseless impulses to buy; frequently buying unneeded items or more than can be afforded; shopping for periods longer than intended; and experiencing adverse consequences, such as marked distress, impaired social or occupational functioning, and/or financial problems.[35] Persons with compulsive buying disorder often report irresistible or uncontrollable impulses to buy or shop; mounting tension or anxiety with the impulses; and relief of tension and/or pleasurable feelings (e.g. 'a high', 'a buzz', or 'a rush') with the act of buying or shopping. The disorder is associated with distress, financial and legal difficulties, and impairment in social and vocational functioning.[34,35]

Epidemiology and course

Compulsive buying behaviour is thought to be common, with an estimated lifetime prevalence of 5.8 per cent in the United States general adult population.[36] Indeed, in a recent study of the prevalence of various impulse control disorders in a psychiatric inpatient population, compulsive buying disorder was the most common current (9.3 per cent) and lifetime (9.3 per cent) impulse control disorder diagnosis.[18]

Compulsive buying disorder is probably more common in women than men.[34,35] It may begin in adolescence or adulthood and usually has either an episodic or a chronic course. The course of compulsive buying disorder into old age is unknown.

Associated psychopathology and comorbidity

Compulsive buying disorder often co-occurs with mood, anxiety, substance use, eating, and other impulse control disorders in clinical samples.[34,35] It may also be associated with certain personality disorders, but this comorbidity has received less systematic attention.[34]

Family studies

Preliminary research, including one controlled study, suggests compulsive buying disorder is associated with increased familial rates of mood, substance use, and possibly impulse control disorders.[34,35]

Psychobiology

In a molecular genetic study, no association was found between two serotonin transporter gene polymorphisms and compulsive buying disorder.[34]

Treatment response

Isolated reports of psychoanalytic and insight-oriented psychotherapy in compulsive buying disorder have mostly been unsuccessful. Cognitive behavioural therapy may hold promise. Two patients with compulsive buying disorder each responding to cue exposure plus response prevention after failing clomipramine treatment have been described.[12] In addition, a randomized study found group cognitive behavioural therapy (N=28) superior to wait-list control (N=11) in reducing compulsive buying episodes, time spent buying, and scores on buying symptom measures in 39 patients with compulsive buying disorder.[37] Support groups patterned after Alcoholics Anonymous, such as Debtors Anonymous, self-help books, and financial counselling are available, but their effectiveness has not been formally evaluated.[12,34]

Various antidepressant medications have been reported to be effective for compulsive buying in case reports and open trials, but two randomized, placebo-controlled, double blind studies of fluvoxamine in a total of 54 patients with compulsive buying (N=17 and N=37, respectively) failed to show separation between drug and placebo.[12,34] Both studies, however, were limited by high placebo response rates, possibly due to the use of diaries to record buying behaviour. In a 7-week, open-label trial of citalopram (N=24) followed by a 9-week double-blind, placebo-controlled continuation trial that omitted use of shopping diaries for the 15 subjects who met responder criteria, none of 7 patients randomized to remain on citalopram relapsed as compared with 5 (63 per cent) of 8 patients randomized to receive placebo (P=0.019)[12,34].

Patients with compulsive buying have also been reported to respond to mood stabilizers, naltrexone (at 100 mg/day but not 50 mg/day), and topiramate.[12,34]

Repetitive self-mutilation

Clinical description

Repetitive self-mutilation, also called impulsive deliberate self-harm, parasuicide, or self-injurious behaviour, is the repeated,

direct destruction of body tissue without suicidal intent.[38,39] Examples include skin cutting, skin burning, self-hitting, severe skin scratching, and even bone breaking. A wide range of body parts are mutilated, such as arms, legs, abdomen, head, chest, and genitals.

Numerous clinical studies suggest that this syndrome often meets the DSM-IV and ICD-10 definitions of impulse control disorders.[38] Specifically, repetitive self-mutilation is characterized by intrusive, recurrent, and irresistible impulses to harm oneself without suicidal intent that are associated with increasing tension, anxiety, anger, or other dysphoric affective states, along with relief of the uncomfortable affect with or shortly after the act of self-harm. In addition, the act of self-harm is often not associated with pain (i.e. associated with analgesia) and performed privately.[38,39]

Epidemiology and course

The prevalence of narrowly defined repetitive self-mutilation is unknown.[38,39] Clinical studies, however, suggest that the condition is more common in females than males, usually begins early in life (e.g. late childhood, adolescence, and early adulthood), and may persist for 10 to 15 years.[38,39]

Associated psychopathology and comorbidity

Repetitive self-mutilation often co-occurs with other Axis I and II psychiatric disorders.[38,39] These include mood, substance use, eating, psychotic, dissociative, and borderline personality disorders. Repetitive self mutilation may also co-occur with suicide attempts and adverse childhood experiences in patients with certain pathologies, especially borderline personality disorder.[39] Of note, although deliberate self injury is a core feature of borderline personality disorder, not all patients with repetitive self mutilation have borderline personality disorder.[38]

Family history

No family history studies of repetitive self-mutilation have been conducted.

Psychobiology

Although studies have consistently found an association between low central CSF 5-H1AA levels with both impulsive aggression and violent suicide, the results of such studies in patients with repetitive self mutilation have been mixed—with some, but not all, finding similar reductions.[38,39] One study found that broadly-defined self harm was associated with allelic variation in the tryptophan hydroxylase gene (TPH A779C), but not with polymorphisms of five other serotonergic genes.[40]

Studies have found increased pain thresholds in borderline personality patients with repetitive self mutilation who are analgesic to the pain of their self injurious behaviour, suggesting dysfunction the endogenous opioid system.[38,39] In support of this possibility, one study found elevated plasma metenkephalin studies in a small group of analgesic self-injuring persons. Another study, however, found pretreatment with naltrexone did not reduce the anesthesia (as evaluated by the cold pressor test) of a similar group of subjects.[39]

Treatment response

There are no controlled treatment studies of narrowly-defined repetitive self-mutilation. However, two psychological treatments, dialectal behaviour therapy and psychoanalytically-oriented partial hospitalization, have each been shown superior to 'treatment as usual' in decreasing chronic parasuicidal behaviour in women with borderline personality disorder.[39] In addition, a 1998 meta-analysis of 20 treatment trials of broadly-defined deliberate self harm indicated significantly reduced repetition of self harm for problem solving therapy and provision of an emergency contract card in addition to standard care.[41]

Regarding medical treatments, agents with antidepressant, mood-stabilizing, antipsychotic, anticonvulsant, and anti-opiate properties have been reported to be effective in case reports or case series.[12,33,39] In the above noted 1998 meta-analysis, a significantly reduced rate of further self harm was observed for depot flupenthixol versus placebo in the one study of antipsychotic medication that was located and analyzed.[41] The two studies of antidepressants evaluated, however, showed no benefit.

Pathological skin picking

Clinical description

Pathological skin picking (also called neurotic or psychogenic excoriation, compulsive skin picking, dermatotillomania, and *acné excorié*) is excessive scratching, picking, gouging, or squeezing of the skin sometimes in response to an itch or other skin sensation or to remove a lesion on the skin.[42,43] Most patients use fingernails to excoriate the skin, but the teeth and instruments (for example, tweezers, nail files, pins, or knives) are also used. Pathological skin picking causes substantial distress in patients, with most experiencing functional impairment and many reporting medical complications, some severe enough to warrant surgery.

Although not recognized as a distinct DSM-IV or ICD-10 disorder, pathological skin picking resembles an impulse control disorder in that patients sometimes experience an increase in tension prior to picking with transient relief or pleasure with picking or immediately afterwards. Many patients find themselves acting automatically. It also has compulsive features, in that it is repetitive, ritualistic, anxiety reducing, often resisted, and egodystonic. Moreover, some patients describe obsessions about an irregularity on the skin or preoccupations with having smooth skin.

Epidemiology and course

Pathological skin picking may occur in about 2 per cent of dermatology clinic patients, predominately in women, and up to 3.8 per cent of college students.[42,43] The disorder may begin in adolescence or adulthood, and the mean duration of symptoms has ranged from 5 to 18 years, with a better prognosis for patients who have had the symptoms for less than 1 year.[42]

Associated psychopathology and comorbidity

Pathological skin picking often co-occurs with mood, anxiety, and somatoform disorders.[42] It is especially common in body dysmorphic disorder.[43] The comorbidity of pathological skin picking and personality disorders has not been systematically studied.

Family studies

Preliminary family history data suggest first-degree relatives of probands with pathological skin picking may have elevated rates of mood and substance use disorders.[42]

Treatment response

Various behavioural treatments (e.g. habit reversal) may be effective in pathological skin picking.[12] One small placebo-controlled trial (N=21) found that fluoxetine (mean dose of 55 mg/day) may be beneficial[12]. Other serotonin reuptake inhibitors, the tricyclic doxepin (which has been hypothesized to have antipruritic properties due to its antihistaminic effects), inositol, the glutamate-modulating agent riluzole, and certain direct skin treatments (dermatologic and surgical) may also be effective.[12,32,42,44,45]

Onychophagia

Clinical description

Onychophagia is repetitive and excessive nail-biting.[46] The cuticles and skin around the nails are also frequently bitten, picked, or clipped. Onychotillomania, the excessive picking, clipping, or tearing of the nail, may be a variant.

Although not classified as a psychiatric disorder in DSM-IV or ICD-10, onychophagia resembles an impulse control disorder in that the behaviour is often irresistible, automatic, and associated with an increase of tension before and relief or pleasure during or immediately after its enactment.[46] It also has compulsive features in that it is repetitive, resisted, and associated with relief of anxiety.

Epidemiology and course

Nail-biting is more common in children than adults, and may affect 5 to 10 per cent of adults over the age of 30 years.[46] Boys and girls are affected equally until after the age of 10 years, when nail-biting becomes more common in boys.

Associated psychopathology and comorbidity

Onychophagia may be associated with mood, anxiety, and substance use disorders. Of 25 adult subjects who underwent a medication trial for onychophagia, 17 (68 per cent) had a lifetime Axis I psychiatric disorder—despite the exclusion of subjects with obsessive–compulsive disorder, a primary major affective disorder, current substance abuse, or psychosis.[46] Four subjects (16 per cent) had at least one personality disorder.

Family studies

Severe nail biting may be familial.[46,47] Of 112 family members of 25 subjects entering a medication trial for onychophagia, seven (6 per cent) had severe nail-damaging behaviour, four (4 per cent) were severe nail-biters, and three (3 per cent) picked or chewed their hands or feet.[46] In addition, twin studies have found higher concordance rates of nail biting in monozygotic compared with dizygotic twins.[47]

Treatment response

Various cognitive behavioural therapies (especially habit reversal, but also self-monitoring, use of bitter tasting substances, competing responses, and negative practice training) are probably effective in onychophagia.[12] In the only controlled pharmacotherapy study of onychophagia, clomipramine (mean dose 120 mg/day) was superior to desipramine (mean dose 135 mg/day) in eliminating nail-biting, reducing nail-biting severity and impairment, and in improving overall clinical progress.[46]

Conclusion

Growing research shows that the impulse control disorders are much more common than once thought to be. The consistency of the 'structure' of the irresistible impulse (a core disturbance of impulsivity and compulsivity) together with increasing research showing that it responds to certain treatments, especially cognitive-behavioural psychotherapies and medical treatments with thymoleptic or anticraving properties, regardless of its 'content' (the specific impulse experienced), strongly suggest that it is an important psychopathological symptom, and that impulse control disorders are legitimate mental disorders that are in fact likely to be related despite their apparent differences.

Further information

Hollander, E. and Stein, D.J. (eds.) (2006). *Clinical Manual of Impulse-Control Disorders.* American Psychiatric Publishing, Inc., Arlington, VA.

Gabbard, G.O. (ed.) (2007). *Gabbard's Treatments of Psychiatric Disorders* (4th ed.). American Psychiatric Publishing, Inc., Arlington, VA.

References

1. Hollander, E. and Stein, D.J, (eds.) (2006). *Clinical manual of impulse control disorders*. American Psychiatric Publishing, Arlington, VA.
2. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Association, Washington, DC.
3. World Health Organization. (1992). *International statistical classifi cation of diseases and related health problems*, 10th revision. WHO, Geneva
4. Coccaro, E.F. and Danehy, M. (2006). Intermittent explosive disorder. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein,), pp. 19–37. American Psychiatric Publishing, Arlington, VA.
5. Kessler, R.C., Coccaro, E.F., Fava, M., *et al.* (2006). The prevalence and correlates of DSM-IV intermittent explosive disorder in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **63**, 669–78.
6. McElroy, S.L., Soutullo, C.A., Beckman, D.A., *et al.* (1998). DSM-IV intermittent explosive disorder: a report of 27 cases. *Journal of Clinical Psychiatry*, **59**, 203–10.
7. Perlis, R.H., Smoller, J.W., Fava, M., *et al.* (2004). The prevalence and clinical correlates of anger attacks during depressive episodes in bipolar disorder. *Journal of Affective Disorders,* **79**, 291–5.
8. Linnoila, M., DeJong, J. and Virkkunen, M. (1989). Family history of alcoholism in violent offenders and impulsive fire setters. *Archives of General Psychiatry*, **46**, 613–16.
9. Mattes, J.A. and Fink, M. (1990). A controlled family study of adopted patients with temper outbursts. *Journal of Nervous and Mental Disease*, **178**, 138–9.
10. Virkkunen, M., DeJong, J., Bartko, J., *et al.* (1989). Psychobiological concomitants of history of suicide attempts among violent offenders and impulsive fire setters. *Archives of General Psychiatry*, **46**, 604–6.
11. Coccaro, E.F., McCloskey, M.S., Fitzgerald, D.A., *et al.* (2007). Amygdala and orbitofrontal reactivity to social threat in individuals with impulsive aggression. Biological Psychiatry, **62,** 168–71
12. McElroy, S.L. and Keck, P.E. Jr (2007). Impulse control disorders. In *Gabbard's treatments of psychiatric disorders,* (4th edn),(ed. G.O. Gabbard), pp. 877–88. American Psychiatric Publishing, Arlington, VA.
13. Goedhard, L.E., Stolker, J.J., Heerdink, E.R. *et al.* (2006). Pharmacotherapy for the treatment of aggressive behavior in general adult psychiatry: a systematic review. *Journal of Clinical Psychiatry,* **67**, 1013–24.

14. McElroy, S.L., Keck, P.E. Jr, Pope, H.G.J., *et al.* (1991). Kleptomania: clinical characteristics and associated psychopathology. *Psychological Medicine*, **21**, 93–108.
15. McElroy, S.L., Pope, H.G. Jr, Hudson, J.I., *et al.* (1991). Kleptomania: a report of 20 cases. *American Journal of Psychiatry*, **148**, 652–7.
16. Bayle, F.J., Caci, H., Millet, B., *et al.* (2003) Psychopathology and comorbidity of psychiatric disorders in patients with kleptomania. *American Journal of Psychiatry*, **160**, 1509–13.
17. Grant, J.E., (2006). Kleptomania. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein), pp. 175–201. American Psychiatric Publishing, Arlington, VA.
18. Grant, J.E., Levine, L., Kim, D., *et al.* (2005). Impulse control disorders in adult psychiatric inpatients. *American Journal of Psychiatry*, **162**, 2184–8.
19. Lejoyeux, M., McLoughlin, M. and Adès, J. (2006). Pyromania. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein), pp. 229–50. American Psychiatric Publishing, Arlington VA.
20. Bienvenu, O.J., Samuels, J.F., Riddle, M.A., *et al.* (2000). The relationship of obsessive-compulsive disorder to possible spectrum disorders: results from a family study. *Biological Psychiatry*, **15**, 287–93.
21. Grant, J.E., Correia, S. and Brennan-Krohn, T. (2006).White matter integrity in kleptomania: A pilot study. Psychiatry Research, **147**, 233–7.
22. Koran, L.M., Aboujaoude, E.N. and Gamel, N.N. (2007). Escitalopram treatment of kleptomania: an open-label trial followed by double-blind discontinuation. *Journal of Clinical Psychiatry*, **68**, 422–7.
23. Lewis, N.D.C. and Yarnell, H. (1951). Pathological firesetting (pyromania). Nervous and mental disease monograph 82. Coolidge Foundation, New York.
24. Grant, J.E. (2006). SPECT imaging and treatment of pyromania. *Journal of Clinical Psychiatry,* **67**, 998.
25. Bourget, D. and Bradford, J. (1987). Fire fetishism, diagnostic and clinical implications: a review of two cases. *Canadian Journal of Psychiatry*, **32**, 459–62.
26. Franklin, M.E., Tolin, D.F. and Diefenbach, G.J. (2006). Trichotillomania. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein), pp. 1149–73. American Psychiatric Publishing, Arlington, VA.
27. Woods, D.W., Flessner, C., Franklin, M.E., *et al.* (2006). Understanding and treating trichotillomania: what we know and what we don't know. *Psychiatric Clinics of North America*, **29**, 487–501.
28. Schlosser, S., Black, D.W., Blum, N., *et al.* (1994). The demography, phenomenology, and family history of 22 persons with compulsive hair pulling. *Annals of Clinical Psychiatry*, **6**, 147–52.
29. Keuthen, N.J., Makris, N., Schlerf, J.E., *et al.* (2007). Evidence for reduced cerebellar volumes in trichotillomania. *Biological Psychiatry,* **61**, 374–81.
30. Zuchner, S., Cuccaro, M.L., Tran-Viet, K.N., *et al.* (2006). SLITRK1 mutations in trichotillomania. *Molecular Psychiatry*, **11**, 887–9.
31. Hemmings, S.M., Kinnear, C.J., Lochner, C., *et al.* (2006). Genetic correlates in trichotillomania – a case-control association study in the South African Caucasian population. *Israel Journal of Psychiatry and Related Sciences*, **43**, 93–101.
32. Lochner, C., Seedat, S., Niehaus, D.J., *et al.* (2006). Topiramate in the treatment of trichotillomania: an open-label pilot study. *International Clinical Psychopharmacology*, **21**, 255–9.
33. Seedat, S., Stein, D.J. and Harvey, B.H. (2001). Inositol in the treatment of trichotillomania and compulsive skin picking. *Journal of Clinical Psychiatry*, **62**, 60–1.
34. Black, D.W. (2006). Compulsive shopping. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein), pp. 203–28. American Psychiatric Publishing, Arlington, VA.
35. McElroy, S.L., Keck, P.E. Jr., Pope, H.F. Jr., *et al.* (1994). Compulsive buying: a report of 20 cases. *Journal of Clinical Psychiatry*, **55**, 242–8.
36. Koran, L.M., Faber, R.J., Aboujaoude, E., *et al.* (2006). Estimated prevalence of compulsive buying behavior in the United States. *American Journal of Psychiatry*, **163**, 1806–12.
37. Mitchell, J.E., Burgard, M., Faber, R., *et al.* (2006). Cognitive behavioral therapy for compulsive buying disorder. *Behavior Research and Therapy*, **44**, 1859–65.
38. Favazza, A.R. (1998). The coming of age of self mutilation. *Journal of Nervous and Mental Disease*, **186**, 259–68.
39. Simeon, D. (2006). Self-injurious behaviors. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein), pp. 63–81. American Psychiatric Publishing, Arlington, VA.
40. Pooley, E.C., Houston, K., Hawton, K., *et al.* (2003). Deliberate selfharm is associated with allelic variation in the tryptophan hydroxylase gene (TPH A779C), but not with polymorphisms in five other serotonergic genes. *Psychological Medicine*, **33**, 775–83.
41. Hawton, K., Arensman, E, Townsend, E., *et al.* (1998). Deliberate self harm: systematic review of effi cacy of psychosocial and pharmacological treatments in preventing repetition. *British Medical Journal*, **327**, 441–7.
42. Arnold, L.M., Auchenbach, M.B. and McElroy, S.L. (2001). Psychogenic excoriation. Clinical features, proposed diagnostic criteria, epidemiology and approaches to treatment. *CNS Drugs*, **15**, 351–9.
43. Grant, J.E., Menard, W. and Phillips, K.A. (2006). Pathological skin picking in individuals with body dysmorphic disorder. *General Hospital Psychiatry*, **28**, 487–93.
44. Sasso, D.A., Kalanithi, P.S., Trueblood, K.V., *et al.* (2006). Beneficial effects of the glutamate-modulating agent riluzole on disordered eating and pathological skin-picking behaviors. *Journal of Clinical Psychopharmacology*, **26**, 685–7.
45. Bowes, L.E. and Alster, T.S. (2004). Treatment of facial scarring and ulceration resulting from ***acné excorié*** with 585-nm pulsed dye laser irradiation and cognitive psychotherapy. *Dermatologic Surgery*, **30**, 934–8.
46. Leonard, H.L., Lenane, M.C., Swedo, S.E., *et al.* (1991). A double-blind comparison of clomipramine and desipramine treatment of severe onychophagia (nail biting). *Archives of General Psychiatry*, **48**, 821–7.
47. Ooki, S. (2005). Genetic and environmental influences on finger-sucking and nail-biting in Japanese twin children. *Twin Research and Human Genetics*, **8**, 320–7.

4.13.2 Special psychiatric problems relating to gambling

Emanuel Moran

Introduction

Gambling is an activity with the following elements:

- A contract between two or more people, which is based on a forecast of the outcome of an uncertain event involving random processes.
- Property, referred to as the stake, is transferred between those taking part, so that some gain at the expense of others.
- The property transfer depends on the outcome or result of the uncertain event, which has been forecast.
- Participation is voluntary and not necessarily related to gaining the property, but used to obtain an experience.

Clinical features

Gambling misuse is a behavioural disorder that can usually be recognized by the presence of any of the following features:

- Excessive gambling either in terms of the money spent or the time devoted.
- Intermittent or continuous preoccupation with gambling and the development of tolerance and craving for it.
- Loss of control over gambling and 'chasing of losses', despite the realization that damage is resulting.
- Disorder affecting the person who is gambling and the family:
  - financial disturbances, such as debt and shortage;
  - social disturbances, such as loss of employment and friends, running away from home, eviction, marital problems, divorce, behaviour disorders in the children of the family, criminality and imprisonment;
  - psychological disturbances, such as depression and attempted suicide.

Classification

In the past, this syndrome has been referred to as compulsive gambling. However, it is not a true obsessive–compulsive state but a heterogeneous group of conditions, characterized by excessive gambling resulting in disturbance for those involved. The term 'pathological gambling' is more appropriate, since it is not based on any assumptions regarding the underlying processes.[1]

ICD-10[2] describes pathological gambling as a form of behaviour under 'habit and impulse disorders'. On the other hand, DSM-IV[3] implies a homogeneous disease entity and provides criteria for its recognition under 'impulse-control disorders not elsewhere classified'. The ICD-10 approach is preferable since it emphasizes the fact that the condition is a behavioural disorder resulting from faulty habits.

Five varieties of pathological gambling can be recognized[4,5]:

- Subcultural gambling arises out of the person's background, which is one of socially accepted heavy gambling.
- Impulsive gambling is characterized by loss of control for varying periods and the tendency to be associated with tolerance, craving, and dependence on the activity.
- Neurotic gambling occurs as a response to an emotional problem, particularly in a disturbed relationship in marriage or during adolescence.
- Symptomatic gambling occurs in mental illness, usually depression, which is the primary disorder.
- Psychopathic gambling is part of the generalized disturbance of behaviour that characterizes antisocial personality disorder.

Diagnosis

For various social reasons, pathological gambling is most easily recognized in men since they tend to patronize those types of gambling that have a high turnover of money so that excess is more likely to become apparent.

Women have tended to gamble on lotteries, bingo, and football pools. These may not involve such large sums of money and excess often presents with disturbances in the social sphere rather than through the accumulation of large debts. However, the greater general acceptance of gambling and the advent of remote gambling via the Internet, television, and mobile devices is changing the situation considerably.

While pathological gambling is seen in all age groups, an increasing number of children and young people are presenting with the condition as a result of gambling on slot/gaming machines. Also, in recent years, remote gambling among children and young people has led to increasing problems. This is in spite of the fact that most jurisdictions treat gambling as an adult activity. Pathological gambling in adulthood frequently has its origins in heavy gambling in childhood and adolescence.

Aetiology and epidemiology

The nature of gambling

The experience of risk provides amusement, thrill, and excitement and is therefore pleasurable. These experiences make gambling attractive and the stake money is used to purchase them, with winnings as an occasional bonus. A few, who gamble professionally, are also able to win money regularly because they have sources of information that reduce the uncertainty, as in betting on horses and dogs. Their gambling is planned and deliberate.

Gambling is usually organized commercially with the odds in favour of the provider. There is therefore an in-built financial disadvantage to those who use the facilities. In slot/gaming machines where the provider is at a distance from the gambling event, this is often not apparent to those who take part.

Commercial gambling involves large sums of money, and has traditionally been confined to licensed premises. Those present have gone there because they have decided to take part in the gambling. However, developments in technology have made it possible to provide gambling facilities on a remote basis via the Internet, television, and mobile devices.

A number of features inherent in the activity of gambling have effects that make it difficult for a person to stop.

Psychological effects

- Underlying all gambling activity is operant conditioning with intermittent variable ratio reinforcement.[6] This is a most effective schedule for habit-formation and produces a stable, persistent response. Consequently, the long-term net gain or loss to those who gamble is almost irrelevant to the continuation of the activity. It is most dramatically seen in slot/gaming machines, which consequently are the main source of profit for the gambling industry.
- Rapid turnover gambling as in casinos restricts the ability of those who gamble to apply any considered judgement. Inevitably, gambling becomes more impulsive, easily leading to excessive participation.
- The assessment of probability of winning (psychological probability) in the gambling situation differs from the mathematical probability. At low probabilities, it is higher than the mathematical probability and at moderate and high probabilities, it is lower. This even occurs in people who are mathematically knowledgeable.

- In a gambling situation involving only random processes, where the outcome of successive events is completely independent, there is usually the irrational belief that a string of losses makes a win more likely. This is the negative recency effect, which is also referred to as the'Monte Carlo Fallacy' since it forms the basis of many spurious gambling systems, especially in roulette. Paradoxically, it is associated with the belief that a string of wins is likely to continue ('a lucky streak'). Also, a 'near win' generally tends to be treated as a prelude to a win. These illogical ways of thinking encourage continuous gambling and are exploited by slot/gaming machines and lottery scratch cards called 'heart stoppers'.
- Large prizes, even at very low probabilities, entice the gambler because of the *possibility* of winning. The stimulant effect of roll-overs in lotteries illustrates this.
- Skill in gambling is usually overrated and often implies an unrealistic ability to control the uncertain event that is the subject of the gamble. Thus, in dog and horse race betting, punters tend to place their bets just before 'the off' in the fantasy that this will affect the result.
- There is a tendency to lose track of time during a gambling session.

These psychological effects have been increased as a result of recent developments in commercial gambling.

- Loyalty cards providing rewards for money spent on gambling in a particular facility.
- Remote gambling on the Internet, television, and mobile devices has resulted in the following:
  - The convenience and anonymity of gambling being available in an isolated domestic setting, without the checks and constraints that can be exercised by the presence of others as in licensed premises.
  - Monetary credit in the form of e-cash systems, reducing the likelihood that those gambling will set a limit on the money staked.
  - Behavioural targeting and messages on some online gambling sites encouraging further gambling when an attempt is made to stop.
  - Difficulty in preventing children and young people from having access, especially since the advent of online social networks.

Physical effects

- A gambling loss in normal subjects immediately results in particular localized activity in the medial frontal cortex of the brain. This is then associated with subsequent *more risky* gambling choices. This is consistent with the negative recency effect.
- Disturbances involving the reward pathways in the brain are significantly associated with excessive gambling.
- There is a great range and strength of emotions during gambling decisions associated with cortical responses in the brain to the expectation of winning money.
- Even normal, social levels of drinking alcohol that alter self-control over decision-making, increase the difficulty in deciding at what point to stop, when losing, in a gambling situation.

Predisposing factors

In the presence of available gambling facilities, certain predispositions may increase the likelihood of pathological gambling.

Morbid risk-taking

Since gambling is a type of risk-taking, it lends itself to be used by those who, for reasons related to their personality, have a high need for risk. They spend large sums of money on the intangible commodity of risk, which may easily pass unnoticed because it is fleeting.

This morbid propensity to take risks shows itself in other ways. Thus, the incidence of attempted suicide is high among those whose gambling is pathological.[5]

Other personality factors

Freud's formulation of gambling was that it resembles masturbation, is a substitute for it, and is resorted to in the context of unresolved Oedipal difficulties.[7] Others have pointed out that pathological gambling may be a manifestation of self-punishment, with an unconscious desire to lose, arising from a psychological mechanism referred to as 'psychic masochism'.[8]

Those whose gambling is pathological appear to have other predisposing personality traits. They view their behaviour as being largely determined by factors outside their personal control. They also tend towards greater impulsivity.[5,9]

Learning processes

Apart from the winnings and losses, the gambling situation itself may affect learning. As far as the random processes inherent in gambling are concerned, all participants, even a total failure, stand on an equal footing. This may be the only circumstance in which some people have this experience. Gambling may therefore provide a means of dealing with morbid anxiety in the presence of feelings of inadequacy, leading to a conditioned avoidance reaction.

Mental disorder

Pathological gambling may occur in any mental disorder. However, it is most commonly associated with depression. More usually, a neurotic type of depression occurs after a bout of heavy gambling with large losses. In symptomatic pathological gambling, the depression is primary and the gambling is a response to the tension and feelings of guilt that occur in depression. This latter situation is similar to alcohol misuse and shoplifting, as part of the depressive syndrome. Pathological gambling may also be a manifestation of antisocial personality disorder.

Misuse of alcohol and pathological gambling can occur together; either may be the primary disorder and either may lead to the other.

Constitutional factors and physical disorder

Twin studies have demonstrated that the likelihood of pathological gambling occurring in a person is influenced, to an important degree, by inherited factors and/or experiences shared during childhood.[10]

There also appears to be a significant association between pathological gambling and genetic abnormalities involving the dopamine reward pathways.[11] Disturbances of serotonergic, noradrenergic, and dopaminergic neurotransmitter systems have

all been implicated in the aetiology of pathological gambling. This is particularly so in relation to the arousal, behavioural initiation, behavioural disinhibition, and reward/reinforcement mechanisms that are evident in this condition.[12]

There have been reports of pathological gambling associated with dopamine agonist administration for Parkinsonism.[13]

Course and prognosis

The natural history of pathological gambling is one characterized by exacerbations and remissions, often related to life events. Important elements in this are relationships within the family, especially with the spouse/partner. An example of this is the not infrequent sequence of an exacerbation of heavy gambling in the husband, at the time of the wife's first pregnancy.

The outlook in pathological gambling is usually determined by the integrity of the underlying personality. In those in whom the condition appears as a symptom of a neurotic disorder or depression, the prognosis depends on that of the underlying disorder.

Management and treatment

Pathological gambling involves a whole way of life, which has many ramifications. If its management is to be successful, there need to be major changes in the lifestyle of the person concerned. It is best dealt with by a team approach involving at least a psychiatrist, psychologist, and social worker and must include the spouse/partner. Recently, counselling services have been set up but their efficacy has yet to be established.

Assessment of the problem

The following aspects are important:

- An appraisal of the extent and amount of present gambling.
- A history of the development of the gambling from its early beginnings, which is best done if the person being assessed provides the information by means of a detailed written narrative.
- A discussion of this written narrative.
- An indication of the person's motivation, since many who seek help for pathological gambling readily admit that they enjoy it and only want assistance for the problems that have resulted.
- At least initially, an immediate period of total abstinence from gambling.

Supervision of the finances

Excessive gambling is usually associated with a disturbed appreciation of the value of money. In view of this and the continued temptation to gamble, the family finances should be dealt with as follows:

- All monies should be controlled, at least for some time, by the spouse/partner or some trusted person.
- Regular income from wages/salaries should be paid into a bank account over which the spouse/partner or trusted person has sole control.
- A detailed statement should be drawn up of all the outstanding debts, as well as an inventory of the income and outgoings of the person-seeking help and his or her family.
- The person whose gambling has been pathological should discuss the matter with all creditors and agree a repayment plan. This should be consistent with the person's regular income and circumstances to avoid a situation where there would be the temptation to gamble in order to maintain repayments. Since debts are often considerable, these may have to continue over many years.
- After a period of abstinence from gambling, the person whose gambling has been pathological needs to become gradually involved in working jointly with whoever controls the finances.

Counselling

On the basis of information obtained during the course of the initial assessment, the following aspects need further consideration:

- The features inherent in gambling that affect people so that they find it difficult to stop should be highlighted and discussed.
- Social relationships of the person whose gambling has become pathological and the spouse/partner should be reviewed, especially if there have been serious marital problems predating the pathological gambling.
- The way spare time is spent, what friends are cultivated, and what interests are pursued should be reviewed. Since incitement to gamble will have occurred in the past within specific settings, arrangements need to be made to avoid these or, at least, to be prepared for them.
- A joint contract to be reviewed regularly spelling out in detail those types of behaviour to be avoided as well as those to be encouraged may be found helpful.

Gamblers Anonymous

This form of self-help for pathological gambling is organized in regular local groups. As well as meetings for those who have a gambling disorder, there are also separate ones for their spouses/partners. Quite apart from the valuable work done in the group setting, Gamblers Anonymous provides a useful means of establishing alternative social contacts from those that were associated with gambling. Indeed, for some people, Gamblers Anonymous may be the vehicle through which all the necessary help can be provided. Even if this is not the case, Gamblers Anonymous still provides a valuable form of support for the individual and the family.

Psychological treatments

A variety of psychological treatments have been advocated but, in general, their long-term efficacy has not been established. A good outcome has been reported after a cognitive behavioural approach.[14] Also, controlled gambling, rather than permanent abstinence, has led to a reported successful outcome after behavioural treatment.[15]

Psychiatric treatments

Specialist treatment from a psychiatrist and/or a psychotherapist for a neurotic disorder or severe depression may be required, if these clearly underlie the pathological gambling.

Prevention

In view of the nature of gambling and the importance of the social causation of pathological gambling, it is vital that it should be seen as an activity that requires moderation. Unfortunately, the recent increasing reliance of governments and states on gambling for revenue purposes is resulting in a vast growth in the availability of gambling facilities and the incitements to participate.

This has been associated with public policies that actively promote gambling and also claim to encourage moderation. The inconsistency in trying to do both inevitably has a harmful effect on any educational attempt to provide a sensible attitude to gambling. It also undermines any help for those whose gambling has become excessive.

Further information

Journal of Gambling Studies. Springer. http://www.springer.com/west/home/social+sciences/sociology?SGWID=4-40440-70-35680327-0 Online version available http://www.springerlink.com/content/1050-5350

International Gambling Studies. Routledge. http://www.informaworld.com/smpp/title~content=t713701604~tab=sample?action=view&db=all

References

1. Moran, E. (1970). Gambling as a form of dependence. *British Journal of Addiction*, **64**, 419–27.
2. World Health Organization. (1992). *International statistical classification of diseases and related health problems*, 10th revision. WHO, Geneva.
3. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
4. Moran, E. (1970). Varieties of pathological gambling. *The British Journal of Psychiatry*, **116**, 593–7.
5. Moran, E. (1970). Clinical and social aspects of risk-taking. *Proceedings of the Royal Society of Medicine*, **63**, 1273–7.
6. Skinner, B.F. (1966). *Science and human behaviour*. Macmillan, New York.
7. Freud, S. (1961). Dostoevsky and parricide. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 21 (ed. J. Strachey), p. 177. Hogarth Press, London.
8. Bergler, E. (1970). *The psychology of gambling*. International Universities Press, New York.
9. Steel, Z. and Blaszczynski, A. (1998). Impulsivity, personality disorders and pathological gambling severity. *Addiction*, **93**, 895–905.
10. Eisen, S.A., Lin, N., Lyons, M.J., *et al.* (1998). Familial influences on gambling behaviour: an analysis of 3359 twin pairs. *Addiction*, **93**, 1375–84.
11. Cummings, D.E. (1998). The molecular genetics of pathological gambling. *CNS Spectrums: International Journal of Neuropsychiatric Medicine*, **3**, 20–37.
12. DeCaria, C.M., Begaz, T., and Hollander, E. (1998). Serotonergenic and noradrenergic function in pathological gambling. *CNS Spectrum: International Journal of Neuropsychiatric Medicine*, **3**, 38–47.
13. O'Sullivan, S.S. and Lees, A.J. (2007). Pathological gambling in Parkinson's disease. *Lancet Neurology*, **6**, 384–6
14. Sylvain, C., Ladouceur, R., and Boisvert, J.M. (1997). Cognitive and behavioural treatment of pathological gambling: a controlled study. *Journal of Consulting and Clinical Psychology*, **65**, 727–32.
15. Blaszczynski, A., McConaghy, N., and Frankova, A. (1991). Control versus abstinence in the treatment of pathological gambling: a two to nine year follow-up. *British Journal of Addiction*, **86**, 299–306.

4.14

Sleep–wake disorders

Contents

4.14.1 Basic aspects of sleep–wake disorders

Gregory Stores

Introduction

A sound working knowledge of the diagnosis, significance, and treatment of sleep disorders is essential in all branches of clinical psychiatry. Unfortunately, however, psychiatrists and psychologists share with other specialties and disciplines an apparently universal neglect of sleep and its disorders in their training. Surveys in the United States and Europe point to the consistently meagre coverage of these topics in their courses at both undergraduate and postgraduate levels.

The following account is an introductory overview of normal sleep, the effects of sleep disturbance, sleep disorders and the risk of failure to recognize them in psychiatric practice, assessment of sleep disturbance, and the various forms of treatment that are available. The aim is to provide a background for the other chapters in this section.

The close links between the field of sleep disorders and psychiatry which make it essential that psychiatrists are familiar with the field are as follows:

- Sleep disturbance is an almost invariable feature and complication of psychiatric disorders from childhood to old age, with the risk of further reducing the individual's capacity to cope with their difficulties (see Table 4.14.1.3 for further details).
- Sleep disturbance can presage psychiatric disorder.
- Some psychotropic medications produce significant sleep disturbance.
- Of importance to liason psychiatry is the fact that many general medical or paediatric disorders disturb sleep sufficiently to contribute to psychological or psychiatric problems.
- Because of lack of familiarity with sleep disorders and their various manifestations, such disorders may well be misinterpreted as primary psychiatric disorders (or, indeed, other clinical conditions) with the result that effective treatments for the sleep disorder are unwittingly withheld (see later).

Some of these points will be amplified in later sections of this chapter.

Basic features of normal sleep

The scientific study of sleep and its disorders is largely confined to the last several decades. Essentially interdisciplinary advances have displaced earlier speculative accounts including those in psychiatry concerning the significance of dreams, for example. They are well described in recent textbooks of sleep disorders medicine (see recommended sources). Only general points are mentioned here, with special reference to psychiatry where possible.

The nature of sleep

Sleep has characteristic physiological features which distinguish it from other states of relative inactivity. Two distinct sleep states have been defined, that is non-rapid eye movement (**NREM**) sleep and rapid eye movement (**REM**) sleep. The onset of sleep is not simply the shutdown of wakefulness but also the switching between wakefulness, NREM and REM sleep involve complicated active neurochemical mechanisms in different parts of the brain.

The functions of sleep

Debate continues about the various theories concerning sleep, each of which has emphasized physical and psychological restoration and recovery, energy conservation, memory consolidation, discharge of emotions, brain growth and various other biological functions including somatic growth and repair, and maintenance of immune

systems. No one theory accounts for all the complexities of sleep and it seems likely that sleep serves multiple purposes.

From the practical point of view, the most obvious observation is that both physical and psychological impairment follows persistent sleep disturbance. Animals totally deprived of sleep for a long periods die with loss of temperature regulation and multiple system failure. As described later, the adverse effects of chronic sleep loss (considered to be common in modern society) on mood, behaviour, and cognitive function can be substantial, with various consequences for personal, social, occupational, educational, and family functioning.

Sleep stages

Conventionally, standard criteria are used to identify different sleep stages according to their characteristic physiological features especially in the electroencephalogram (**EEG**), electrooculogram (**EOG**), and electromyogram (**EMG**).

NREM sleep is divided into four stages of increasing depth. **Stage I** occurs at sleep onset or following arousal from another stage of sleep. This stage represents 4–16 per cent of the main sleep period. **Stage II** contains more slow EEG activity but is still relatively light sleep. It accounts for 45–55 per cent of overnight sleep. **Stage III** (4–6 per cent of total sleep time) contains yet more slow EEG activity. **Stage IV** is characterized by the slowest activity and constitutes 12–15 per cent of sleep. The combination of stages III and IV is called slow wave sleep (**SWS**) or delta sleep and is considered to be the deepest form of sleep from which awakening is particularly difficult. The arousal disorders such as sleepwalking arise from SWS.

REM sleep is physiologically very different. Brain metabolism is highest in this stage of sleep. Spontaneous rapid eye movements are seen and the skeletal musculature is effectively paralysed. Heart rate, blood pressure, and respiration are all variable, body temperature regulation ceases temporarily, and penile and clitoral tumescence occurs. REM sleep usually takes up 20–25 per cent of total sleep time. Most dreams, including nightmares, occur in REM sleep.

Sleep architecture

NREM and REM sleep alternate cyclically throughout the night starting with NREM sleep lasting about 80 min followed by about 10 min of REM sleep. This 90 min sleep cycle is repeated three to six times each night. Each REM period typically ends with a brief arousal or transition into light NREM sleep.

In successive cycles the amount of NREM sleep decreases and the amount of REM sleep increases. SWS is usually confined to the first two sleep cycles. The diagrammatic representation of overnight sleep is known as a **hypnogram**, a simplified form of which is shown in Fig. 4.14.1.1.

In addition to this conventional sleep staging, there has been increasing interest in the microstructural fragmentation of sleep by frequent, brief arousals (seen mainly in the EEG) lasting a matter of seconds without obvious clinical accompaniments. This subtle type of sleep disruption, overlooked by conventional sleep staging, is increasingly associated with impairment of daytime performance, mood, and behaviour.

Circadian sleep–wake rhythms

The timing of sleep (but not its amount) is regulated by a circadian 'clock' in the suprachiasmatic nucleus (**SCN**) of the hypothalamus.

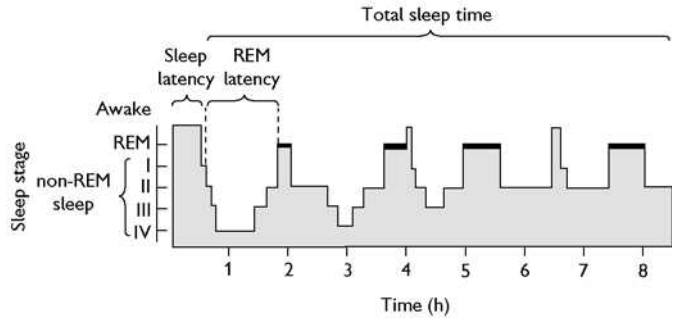


Fig. 4.14.1.1 Diagram of an overnight hypnogram in a young adult.

The intrinsic circadian sleep–wake rhythm is close to 24 h in human adults. Other species are different, an extreme example being dolphins and some other creatures which shut down one cerebral hemisphere at a time ('unihemispheric sleep'), allowing them to be constantly alert. From an early age the individual sleep–wake rhythm has to synchronize with the 24-h day–night rhythm. The main *zeitgeber* by which this is achieved ('entrainment') is sunlight but social cues, such as mealtimes and social activities, are also important.

The SCN also controls other biological rhythms including body temperature and cortisol production with which the sleep–wake rhythm is normally synchronized. In contrast, growth hormone is locked to the sleep–wake cycle and is released with the onset of SWS, whatever its timing.

Melatonin is related to the light–dark cycle rather than the sleep–wake cycle. It is secreted by the pineal gland during darkness and suppressed by exposure to bright light ('the hormone of darkness'). It influences circadian rhythms via the SCN pacemaker which in turn, regulates melatonin secretion by relaying light information to the pineal gland. The widespread popularity of melatonin as a sleep-promoting agent is not justified by what little is known about its action and clinical effectiveness.

Changes with age

Changes in basic aspects of sleep are prominent from birth to old age, although individual differences are seen at all ages. Changes of clinical significance include the following:

- **Total sleep time** decreases with age. Average daily values are as follows: newborns 6–18 h; young children 10 h; adolescents 9 h (although often they obtain significantly less than this); adults 7.5–8 h, including possibly the same in elderly people. The total amount of sleep includes daytime napping in children up to the age of about 3 years.
- **SWS** is particularly prominent in prepubertal children who sleep very soundly. Its decline begins in early adolescence and continues throughout childhood.
- The proportion of **REM sleep** declines from 50 per cent or more of total sleep time in the newborn (more than this in premature babies) to 20–25 per cent by 2 years. This figure remains fairly constant throughout the rest of life. The high level of REM sleep in very early life suggests a role in cerebral maturation but the reason for its persistence throughout life is unclear. Memory processing appears to depend on sleep. However, people deprived

of REM sleep, experimentally or pathologically, can be relatively unaffected either emotionally or cognitively. Deep sleep decreases in the elderly.

- **NREM–REM sleep cycles** occur at intervals of 50–60 min in infants who often enter REM at the start of their sleep period. This interval between sleep cycles remains until adolescence when the periodicity changes to 90–100 min, which persists into adult life. The amounts of NREM and REM in each sleep cycle is about equal in early infancy. Afterwards, NREM sleep (especially SWS) predominates in the earlier cycles and REM sleep in the later cycles.
- **Continuity of sleep** is greatest in pre-pubertal children (as mentioned previously) and least at the extremes of age. Infants are easily awakened and so are the elderly who also wake spontaneously more often. Fragmentation of sleep by brief arousals, or very brief awakenings, is particularly common in old age.
- **Circadian sleep–wake rhythms** change considerably in early development. Full-term neomates show 3–4-h sleep–wake cycles. Sleep periods have largely shifted to the night and wakefulness to daytime by 12 months, except for napping which gradually diminishes and has usually stopped by about 3 years of age. However, a physiological tendency towards an afternoon nap remains throughout the rest of the life. Although repeated brief waking at night is more common in infancy and early childhood than later, it remains a normal occurrence throughout life, increasing in frequency again in old age. The clinical problem arises when there is difficulty returning to sleep after such awakenings.

Psychological effects of sleep disturbance

There is extensive clinical and experimental evidence that sustained sleep disturbance can have serious adverse psychological effects.[1,2] The term sleep disturbance covers the following:

- Loss of sleep (i.e. shortened duration).
- Impaired quality of sleep (repeated disruption of sleep architecture).
- Inappropriate timing of the sleep period in relation to day–night rhythms (as in the various circadian sleep rhythm disorders such as jet lag, shift work, or the more frequently encountered forms seen in clinical practice, as discussed later).

Experimental studies of **total sleep loss** demonstrate a progressive deterioration in cognitive function, mood, and behaviour related to length of sleep loss. However, inter- and also intra-individual differences in susceptibility are seen, reflecting such factors as motivation, personality, and usual sleep requirements. Task characteristics (e.g. brief or prolonged and monotonous tasks), timing of the task in relation to the circadian sleep–wake rhythm, and physical environmental factors such as noise and other distracting stimuli, are also important.

Variations for similar reasons are important in **partial sleep deprivation** experiments which (like those concerning fragmentation of sleep) are much closer to real-life sleep disturbance caused by social activities, job demands, and other aspects of modern lifestyle. These studies raise the issues of how much sleep is needed for optimal daytime functioning and whether these requirements are not being met. It has been argued that there is 'national sleep debt' in the United States and other western countries, and that by sleeping longer than they do habitually, many people would increase their performance and improve their well-being during the day.

The usual subjective effects of sleep disturbance are irritability, fatigue, poor concentration, and depression. More dramatic effects are described with prolonged and severe sleep disturbance, such as disorientation, illusions, hallucinations, persecutory ideas, and inappropriate behaviour with impaired awareness ('automatic behaviour') caused by frequent microsleeps. Psychometric studies have shown that sleep disturbance can produce a range of cognitive impairments, again depending on its duration and individual susceptibility. Sustained attention (vigilance) is particularly vulnerable and possibility abstract thinking and divergent intelligence or creativity.

The experimental findings are in keeping with the results from studies of various occupational groups including junior hospital doctors and drivers of various types of vehicle, in which reduced performance or accidents are associated with sleep disturbance. The common and increasing practice of night-shift work (as part of the '24 h society') is contrary to the fundamental biorhythm of sleeping at night and being awake during the day, and is often accompanied by a reduction in total sleep time and poor quality sleep. It is not surprising that working shifts commonly results in loss of well-being, physical complaints, and impaired productivity and safety, as well as physical disorders. Similarly, the distribution over the 24 h period of road accidents (especially those not involving other vehicles) and other mishaps at work, correspond to that of the levels of sleep tendency assessed objectively. Even industrial and engineering disasters have been attributed to sleep loss and impaired performance on the part of key personnel.

Additional evidence that sleep disturbance affects daytime function comes from neuropsychological studies of certain sleep disorders. Impaired performance on prolonged and complex tasks of subjects with narcolepsy has been shown to be secondary to the effects of their daytime sleepiness rather than an intrinsic neurological deficit. In the many adult patients with obstructive sleep apnoea, attention memory impairment (like depression and irritability commonly reported by these patients) are also largely attributable to daytime sleepiness. There is some evidence that deficits in more complicated 'executive functions' (formulating goals, planning, and carrying out plans effectively) are not necessarily reversed when their sleepiness is relieved by treatment. This might be the result of irreversible anoxic brain changes in the later stages of the condition. Clearly, early detection and treatment of this condition is essential to prevent this happening.

When return to normal sleep is possible, recovery from short periods of sleep disturbance occurs after much less sleep than that originally lost, for example, after one night's sleep following sleep loss over several days and nights. Reversal of the effects of long sleep disturbance in real-life is likely to be complicated, for example by emotional consequences of the disturbance.

Many of the above observations about the psychological effects of sleep disturbance (and their reversibility) have been made on young adult subjects or patients. The area is largely unexplored in other age groups but there is no reason why the general principle should not apply to children and the elderly including demented patients in whom sleep disturbance is particularly prominent.

Another group on whom further research is particularly required are people with learning disabilities (intellectual disability).

The available literature provides good reason to believe that the sleep disorders, especially in the more severely disabled groups, not only affects the majority but also are unusually severe and long-lasting because of lack of appropriate advice and treatment. The sleep disturbance is a problem in its own right and is often associated with various cognitive and behavioural abnormalities which might, at least partly, be the consequence of the sleep disturbance. Sleep disturbance in the duration or the quality of sleep may be one of the few ways of improving to some extent the psychological well-being of people with learning disabilities or dementia (and that of their carers) whose basic condition itself cannot be improved. In the case of the learning disabled, contrary to the common supposition by both professionals and relatives, success can usually be achieved (even in severe and long-standing problems), given an accurate diagnosis of the type of sleep disorder which may be predominantly behavioural or physical in type depending on the cause of the learning disability.

Sleep disturbance in the aetiology of psychiatric illness

Various 'psychotic' and other abnormal psychological phenomena were mentioned earlier resulting from prolonged and severe sleep disturbance, but these are reversed when normal sleep is restored. It remains an open question how often sleep disturbance is a primary cause of psychiatric illness. Evidence is patchy, tentative, and still in need of clarification.

- Over a wide age range, patients with a prior history of insomnia have been found consistently to be at significantly increased risk for severe depression. This could be interpreted in different ways including that sleep disturbance and the depression have a common underlying pathology, or that the sleep problems are an early sign of depression.
- A less fundamental role (but again implying preventative possibilities) is the suggestion that sleep deprivation late in pregnancy and in labour and childbirth at night might trigger post-natal depression.
- Abnormal circadian sleep–wake rhythms have been implicated in various depressive disorders including seasonal affective disorder (SAD). Light therapy has been used to correct the abnormality and relieve the depression and other symptoms.
- Disordered REM sleep mechanisms have (questionably) been considered as fundamental in the development of post-traumatic stress disorder symptoms.
- Some forms of attention-deficit hyperactivity disorder in children are attributed to persistent sleep disturbance.
- In a proportion of patients with schizophrenia, narcolepsy has been reported as the cause of their psychotic symptoms.

Disorders of sleep

Sleep complaints

The starting point for the clinician is the patient's sleep complaint. They are of three basic types:

- Not enough sleep, or unrefreshing sleep (**insomnia**).
- Sleeping too much (**excessive daytime sleepiness** or **hypersomnia**).
- Disturbed episodes during or otherwise related to sleep (**parasomnias**).

The detailed accounts later in this section are organized in relation to these main types of sleep complaint: insomnias (Chapter 4.14.2); excessive daytime sleepiness (Chapter 4.14.3); and parasomnias (Chapter 4.14.4). Sleep problems in childhood and adolescence are discussed in Chapter 9.2.9.

Whatever the clinical setting in which sleep complaints are investigated, the essential aim is to identify the specific sleep disorder from the many other conditions that can give rise to such complaints. Some sleep disorders may cause more than one type of complaint, and a patient may have more than one sleep disorder. The question arises how best to classify the many sleep disorders that have been described.

International classification of sleep disorders—second edition 2005 (ICSD-2)

This system, derived from wide international consultation, is the latest attempt to organize rationally the many ways in which sleep can be disturbed. ICSD-2 replaced the ICSD-Revised scheme outlined in the first edition of this textbook. More than 90 different sleep disorders are grouped as shown in Table 4.14.1.1. The grouping reflects the fact that knowledge about individual sleep disorders is very varied. The basic pathophysiology of some is quite well documented; in others little is known beyond their manifestations, and even they are subject to change as clinical observations improve. As a result, the ICSD-2 groupings are a mixture of those based on a common complaint (e.g. insomnia or hypersomnia), others on presumed aetiology (circadian rhythm sleep–wake disorders), and yet others are grouped according to the organ system from which the problems arise (such as sleep-related breathing disorders). Two additional groups in the system reflect current uncertainty about their status as disorders, or about their true nature.

Each sleep disorder is described in a standardized fashion using a series of sub-headings which include clinical features, demographics, pathology, and differential diagnosis. Treatments are not covered. Some publications are recommended concerning each sleep disorder. An attempt has been made throughout ICSD-2 to highlight aspects of sleep disorders of particular relevance to children.

In all, ICSD-2 provides a concise, easily accessible and up-to-date source of information for consultation by the clinician. Without being over-technical, it is more comprehensive and informative than current ICSD-10 and DSM-IV systems.

Sleep disorders mistaken for primarily psychological or psychiatric conditions

Sleep disorders manifest themselves in many ways. Failure to realize this can result in the misdiagnosis of primary sleep disorders as other types of clinical conditions of psychiatric, neurological, or otherwise medical conditions especially if there is limited familiarity with the sleep disorders field (which, unfortunately, is generally the case). Clearly such mistakes compromise patient care. The following are some of the main examples of this problem. Further details, with examples, are available elsewhere.[3]

Table 4.14.1.1 ICSD-2 groups of sleep disorders

I *Insomnias* The many psychological and physical causes of difficulty getting off to sleep, not staying asleep, early morning wakening, and feeling un-refreshed by sleep are included here including stress, poor sleep habits, and various mental and medical conditions.
II *Sleep-related breathing disorders* This group includes the common condition of obstructive sleep apnoea in adults and in children which often causes daytime sleepiness and other serious effects including changes in mood and behaviour. Central apnoea and various types of hypoventilation/hypoxaemic syndromes are also part of this group.
III *Hypersomnias of central origin not due to a circadian rhythm sleep disorder, sleep-related breathing disorder, or other cause of disturbed nocturnal sleep* Included here are narcolepsy and the causes of intermittent or recurrent hypersomnia such as the Kleine–Levin syndrome.
IV *Circadian rhythm sleep disorders* These disorders are characterized by a mistiming (and often disruption) of the sleep period, resulting in insomnia and/or hypersomnia. A prominent example is the delayed sleep phase syndrome common in adolescence. The advanced sleep phase syndrome can be seen in the elderly in which the sleep period begins in the evening with waking early when sleep requirements have been met. Irregular sleep–wake rhythms may be the result of an ill-organized way of life and substance abuse. Jet lag and nightshift work disorder are further examples of sleep problems caused by disturbance of the biological clock controlling the sleep–wake cycle.
V *Parasomnias* These are abnormal behaviours or sensations during or otherwise closely related to sleep. Many can be categorized according to the stage of sleep with which they are usually associated for example NREM sleep (sleepwalking) and REM sleep (nightmares, REM sleep behaviour disorder). Other parasomnias of particular psychiatric interest include sleep-related dissociative disorders and sleep-related eating disorders.
VI Sleep-related movement disorders These include the restless leg syndrome and periodic limb movement disorder.
VII Isolated symptoms, apparently normal variants and unresolved issues
VIII Other sleep disorders The classificatory scheme also includes an appendix on sleep disorders associated with conditions classifiable elsewhere such as sleep-related epilepsy, headaches, gastro-oesophageal reflux. A further appendix is concerned with other psychiatric and behavioural disorders frequently encountered in the differential diagnosis of sleep disorders. This appendix covers mood disorders, anxiety disorders, somatoform disorders, schizophrenia and other psychiatric disorders, personality disorders, and disorders of a psychiatric or behavioural type first diagnosed in infancy, childhood, or adolescence.

- Persistently not obtaining enough sleep, or having poor quality sleep because of interruptions by frequent subclinical arousals (as in obstructive sleep apnoea), is likely to cause tiredness, fatigue, irritability, poor concentration, impaired performance (possibly causing injuries or accidents at work or while driving), or depression. Out of the various possible explanations for such changes of behaviour, sleep disturbance may well be overlooked with failure to appreciate that, with improvement in sleep (which is usually possible with the right advice), such problems can be resolved. Occupational groups at special risk of sleep disturbance and its harmful effects include some clinicians.
- Excessive sleepiness, whatever its cause out of the many possibilities including physical conditions, is often misjudged as laziness, disinterest, daydreaming, lack of motivation, depression, intellectual inadequacy, or a number of other unwelcome states of mind. Sometimes, in very sleepy states, periods of 'automatic behaviour' occur, i.e. prolonged, complex, and possibly inappropriate behaviour with impaired awareness of events and, therefore, amnesia for them. Such episodes, the result of repeated 'microsleeps', can easily be misconstrued as reprehensible or dissociative features, misbehaviour, or prolonged seizure states. The paradoxical effect in young children of sleepiness causing over-activity has sometimes led to a diagnosis of attention-deficit hyperactivity disorder (ADHD) inappropriately treated with stimulant drugs instead of treatment for the sleep disorder.

A number of specific sleep disorders are at particular risk of being misinterpreted.

- In so-called delayed sleep phase syndrome, in which there is difficulty getting to sleep until very late, and problems getting up in the morning because of a shift in this timing of the sleep phase, is considered to be particularly common in adolescents who may be mistakenly thought to be awkward, lazy, irresponsible, or indulging in school refusal of the more usual type. In fact, this sleep disorder at that age is the result of a combination of normal pubertal biological body clock changes and alterations in lifestyle involving staying up late for social reasons or for study.
- It is not generally appreciated that even very complicated behaviour is possible whilst a person is still asleep as in sleepwalking episodes. Those with agitated sleepwalking or sleep terrors may appear to be very fearful and distressed and rush about and cry out as if escaping from danger. Other sleepwalkers develop an eating disorder with excessive weight gain due to the amount of food they consume while they are still asleep at night. Yet others behave in an aggressive or destructive way causing injury to themselves or other people and, at times sexual or other serious offences have been committed during a sleepwalking episode. If it is not known that such complicated actions are compatible with still being asleep, it is likely to be assumed that the person was awake at the time and aware of what he or she was doing, and, therefore, responsible for what had happened.
- Obstructive sleep apnoea is another case in point where this essentially physical disorder may be mistaken for being something very different from its true nature. The impairment of sleep quality, which characterizes this condition, can cause excessive daytime sleepiness, changes of personality, as well as adverse affects on social life and performance at work, as well as intellectual deterioration to the extent that dementia is suspected.
- Narcolepsy/cataplexy is also at serious risk of being misdiagnosed, sometimes for many years. Neurosis or depression is commonly mistaken for the narcolepsy symptoms.
- In REM sleep behaviour disorder there is a pathological retention of muscle tone during REM sleep so that a person can act out their dreams and behave violently if they have violent dreams. The dramatic behaviour that may result, including attacks on the sleeping partner, is easily misconstrued as intentional aggression.

- The complicated behaviour (far removed from that seen in other seizure states) that characterizes nocturnal frontal lobe epilepsy is often mistakenly thought to be evidence of a psychiatric disorder.
- In addition, sleep paralysis may be misconstrued as a psychotic disorder when (not uncommonly) accompanied by hallucinatory experiences.

In addition to sleep disorders being mistaken for psychiatric disorder, the opposite problem arises occasionally, that is some patients simulate excessive daytime sleepiness in order to avoid a psychologically troubling situation. Similarly, apparent parasomnias during sleep have sometimes been shown by polysomnography to occur when the patient is actually awake.

Detection and assessment of sleep disorders

As suggested earlier, evidence of a sleep disturbance should be actively sought in members of the general population and the various groups, including psychiatric patients, who are at special risk of sleep disorders. Otherwise, many instances of even severe sleep disturbance will continue to go unrecognized and untreated.

Sleep history

Routinely, all patients should be asked the following screening questions:

- Do you sleep long enough or well enough?
- Are you very sleepy during the day?
- Do you do unusual things or have strange experiences at night?

Ideally, their partner or other relatives should also be asked the same questions about the patient because the existence or severity of some forms of sleep disturbance are not known to the patient. In the case of children, parents are the main source of information but teachers' observations about daytime sleepiness or disturbance are also important.

If the answer to any of these enquiries is positive, a detailed sleep history is required. As traditional clinical history-taking schedules pay little attention to sleep, additional sleep-related enquiries will need to be made, covering the following points about the sleep problem.

- Precise nature of the sleep complaint, its onset, development, and current patterns.
- Medical or psychological factors at the onset of the sleep problem or which might have maintained it.
- Patterns of occurrence of the symptoms that is factors making them better or worse, weekdays compared to weekends, or work compared with holiday periods.
- Effect on mood, behaviour, work, social life, other family members.
- Past and present treatments for sleep problems and their effects.
- Past and present medication or other treatments for other illness or disorder.

In addition, detailed information is needed concerning the following:

- The patient's typical 24-h sleep–wake schedule. This can usually start with the evening meal, followed by preparation for and timing of bedtime, time and process of getting to sleep, events during the night, time and ease of waking up and getting up, level of alertness, and mental state and behaviour during the day.
- An attempt should be made to establish the duration, continuity, and timing of the patient's overnight sleep as these are the most important aspects of sleep for daytime functioning. It is also important to identify events of particular diagnostic significance, for example loud snoring.
- Sleep hygiene.

Compilation of a sleep history can be aided by the use of a preliminary sleep questionnaire (e.g. ref.[4]).

Sleep diary

Systematic recording in a booklet each day over 2 weeks or more, using a standardized format, avoids the bias or distortion of retrospective generalizations.

Other histories

Medical and psychiatric histories should include past and current treatment details (in view of the wide range of illnesses or disorders and their treatment with which sleep disturbance is associated). **Social history** should include occupational, marital, and recreational factors (drinking, smoking, drug use), which may affect sleep. **Family history** may be positive, for example in sleepwalking and associated arousal disorders.

Review of systems

Breathing difficulties and nocturia, for example, are associated with sleep disturbance. Severe obstructive sleep apnoea can cause cardio-respiratory and other cardiovascular complications.

Physical and mental state examination

Particular attention should be paid to the following:

- Evidence of any systemic illness including cardio-respiratory disease or neurological disorder (such as Parkinson's disease or stroke) which may disturb sleep.
- Obesity, oral and pharyngeal abnormalities, retrognathia, or mid-face deformity (predisposing to upper airway obstruction).
- Depression or other psychiatric disorder.
- Intellectual impairment, especially features of intellectual disability or dementia (including specific retardation syndromes) in view of their strong association with sleep disturbance.

Audio–video recording and actigraphy

Audio–video recordings can be very instructive in the parasomnias, sometimes revealing a very different picture than that provided in the clinic. Home video systems can be used where admission to hospital is not feasible. Similarly, monitoring of body movements via means of wrist-watch-like devices (**actigraphy**) can be used at home, if necessary, to quantify basic circadian sleep–wake patterns.

Polysomnography (PSG)

Physiological sleep studies are necessary for diagnosis in only the minority of sleep disorders. Traditionally this has entailed admission to a sleep laboratory. However, especially where such facilities are difficult to obtain, where the laboratory situation is unacceptable to patients (including some children or patients who are psychiatrically disturbed), **home PSG**, using portable systems, is useful, although the procedure has yet to be fully standardized and for some disorders is best seen as only a screening procedure. Recording in the home environment has the further advantage that, if the patient is allowed to adapt to the recording procedure before bedtime, the results from a single night's recording can be representative of the patient's habitual sleep. In contrast 'first night effects' are prominent in laboratory recordings and more than one night of polysomnography is required to allow adaptation to take place.

Basic polysomnography entails the recording of EEG, EOG, and EMG. This allows sleep to be staged and a hypnogram to be compiled as illustrated in Fig. 4.14.1.1. Usually the recording is made overnight but it may be continued during the day if required. Basic measures obtained from this information are as follows: total sleep time, time awake and number of awakenings, amount of REM and NREM sleep and their distribution overnight.

PSG can be extended to additional physiological measures, especially the following:

- Respiratory variables and audio–video recordings for sleep-related breathing problems.
- Additional EEG channels (combined with video) if nocturnal epilepsy is suspected.
- Anterior tibialis EMG for the detection of period limb movements in sleep.

Main indications for PSG are:

- The investigation of excessive daytime sleepiness, including the diagnosis of sleep apnoea, narcolepsy, or PLMS.
- The diagnosis of parasomnias where their nature is unclear from the clinical details, where PSG findings contribute essentially to the diagnosis (e.g. REM sleep behaviour disorder), where the possibility of epilepsy exists, or whether there may be more than one parasomnia present.
- As an objective check on the accuracy of the sleep complaint, or response to treatment.

Other investigations

Further laboratory investigations may be appropriate depending on the nature of the sleep problem and the purpose of the assessment.

- **A multiple sleep latency test** (**MSLT**), involving the recording of basic PSG variables, quantifies the degree of daytime sleepiness by measuring the time a patient takes to fall asleep during five opportunities to do so during the day. In adults, a mean sleep latency of 5 min or less indicates pathological sleepiness (usually of organic origin); 5 to 10 min is a grey area which usually includes excessive sleepiness associated with primary psychiatric disorder; while longer than 10 min is normal. These values do not apply in children whose sleep tendency varies with age.
- **HLA typing**, and possibly **CSF hypocretin (orexin) levels** for the investigation of narcolepsy, and **nocturnal penile tumescence monitoring** in the differential diagnosis of organic versus psychogenic impotence, are examples of other specific tests that may be appropriate depending on the clinical problem.

Treatment approaches for sleep disorders

In clinical practice, the pharmacological approach to treatment of sleep problems (especially insomnia) is generally overemphasized, especially in the use of hypnotic-sedative drugs. Table 4.14.1.2 provides some indication of the wide range of available types of treatment, as well as general principles of management, for adults and children. They are roughly arranged in order of the frequency of their use in the comprehensive management of sleep disorders in general. An appropriate choice from this range requires an accurate diagnosis of the underlying sleep disorder. Further details are provided in later contributions to this section on sleep disorders. Claims for the effectiveness of these various measures are based on widespread clinical experience and reports. Few randomized controlled trials have not been published, as yet.

Certain aspects of treatment with special relevance to psychiatric practice are included in Table 4.14.1.3.

Clinical sleep disorders in psychiatric conditions

As mentioned earlier, sleep disturbance is a feature of many psychiatric disorders at all ages. Sometimes a sleep disturbance is profound and constitutes one of the defining characteristics of the psychiatric condition. This aspect of the psychiatric state requires careful definition and quite possibly treatment in its own right, alongside primarily psychiatric help, in order to facilitate recovery.

Table 4.14.1.3 outlines the main types of sleep problem or disorder associated with various psychiatric conditions. Emphasis has been placed on clinical sleep problems rather than PSG abnormalities which are seen in some of the conditions which (although interesting and of potential pathophysiological significance) are

Table 4.14.1.2 Examples of treatment approaches for sleep disorders in adults

General principles
Explain the problem, reassure where appropriate and provide support
Encourage good sleep hygiene (see Chapter 4.14.2)
Where possible treat the underlying cause of the sleep disturbance (e.g. medical or psychiatric disorder)
Take safety or protective measures (e.g. for hazardous parasomnias)
Specific behavioural treatments for insomnia (Chapter 4.14.2)
Chronotherapy (for circadian sleep-wake rhythm disorders)
Sleep phase retiming
Light therapy
Medication
Hypnotics (very selective and short-term)
Stimulants (narcolepsy)
Melatonin (for some circadian sleep–wake rhythm disorders)
Physical measures (e.g. for obstructive sleep apnoea)
Continuous positive airway pressure (CPAP)
Surgery (selected cases)

Table 4.14.1.3 Clinical sleep disturbance in psychiatric disorders

Psychiatric condition	Likely/possible sleep problem/disorder	Possible treatment issues/principles (see also text for general points)
Depression		*General* Emphasis on treatment of depression Possible sleep complications of ADs: drowsiness, insomnia (including SSRIs), RLS, PLMS, RBD
	Insomnia	Early treatment required to prevent worsening of depression Sedating ADs may be helpful
	EDS (minority)	More stimulating ADs may be appropriate Light therapy for SAD
Mania	Profound insomnia	Vigorous treatment required Mood-stimulating drugs may cause arousal disorders
Anxiety disorders	Insomnia and disrupted (poor quality) sleep PTSD nightmares Nocturnal panic attacks	In addition to behavioural treatments; larger dose of antipsychotics at bedtime may be helpful
Eating disorders	Insomnia (especially anorexia nervosa) EDS (especially bulimia nervosa) Sleepwalking with eating behaviours	General principles for insomnia (see Chapter 4.14.2) Possibly SSRIs See Chapter 4.14.4 for management of sleepwalking
Schizophrenia	Insomnia	Most neuroleptics promote sleep As in other psychotic states, behavioural sleep treatments and sleep hygiene are important
Alcohol and other substance abuse (including withdrawal states)	Insomnia including poor qualitysleep Disrupted sleep-wake cycle	Avoid sedative-hypnotic drugs Sleep hygiene principles (Chapter 4.14) and chronotherapy for sleep–wake cycle abnormalities
Alzheimer's disease and other dementing disorders	Progressive insomnia with fragmented sleep and sleep–wake cycle disorders including nocturnal agitated wandering	General principles for insomnia treatments; Sleep problems with some anti-dementia drugs Low-dose antipsychotics, if necessary Promote day–night cues including regular experience of daylight; discourage daytime napping
	RBD	Clonazepam
Child and adult ADHD	Insomnia	Stimulants can add to the sleep problem

KEY:

ADs = Antidepressants
ADHD = Attention-deficit hyperactivity disorder
EDS = Excessive daytime sleepiness
PLMS = Periodic limb movements in sleep
RBD = REM sleep behaviour disorder
RLS = Restless legs syndrome
SAD = Seasonal affective disorder
SSRI = Selective serotonin reuptake inhibitors

not necessarily accompanied by clinical manifestations. A prime example of this is the reduced time between onset of sleep and the start of the first REM sleep period ('REM latency') in certain forms of severe depression. This can be viewed as a type of biological marker which may persist even after the depression itself has lifted, and which may also be seen in relatives of a depressed person with this PSG finding without they themselves suffering from depression. However, the basic significance of reduced REM latency is, as yet, obscure and not helped by the fact that it has also been described in other psychiatric conditions, narcolepsy, following withdrawal from REM sleep-suppressing substances, during recovery from sleep deprivation, and a number of other diverse circumstances.

Another aspect of the pathophysiological relationship between sleep and mood which awaits clarification is a paradox that, while depression is usually associated with sleep disturbance and loss, sleep deprivation is reported to have an antidepressant effect in some patients although the effect does not persist beyond the period of depression.

Clinical sleep disorders associated with neurological and other medical illness

Of obvious relevance to psychiatric practice in general, but liaison psychiatry in particular, is the fact that disturbed sleep (often severe) is associated with many medical conditions. Main examples of this are shown in Table 4.14.1.4 together with mention of treatment issues in relation to each type of illness. Additional general considerations regarding treatment are as follows:

- The main emphasis in the management of sleep disorders in this context should be placed on the treatment of the medical

Table 4.14.1.4 Clinical sleep disturbance in neurological and other medical illness

Medical condition	Likely/possible sleep problem/disorder	Possible treatment issues/principles (see also text for general points)
Parkinson's disease and/or related syndromes	Progressive insomnia including poor quality sleep	Mainly behavioural treatment and sleep hygiene measures as in other insomnias (see Chapter 4.14.2)
	Parasomnias (e.g. nightmares, nocturnal hallucinations)	Consider possible effects of anti-Parkinson's medication on these and other sleep problems
	EDS including sleep attacks	Possibly daytime stimulants
	PLMS	Usual treatments for PLMS, if severe
	RBD	Clonazepam usually effective
Epilepsy	Disrupted sleep depending on type, cause, and severity	Sleep generally improves with seizure control
	EDS	Sedating AEDs can be a factor
Stroke	Insomnia	Usual measures for insomnias
	OSA	
Head injury	Effects depend on type and severity	
	Poor quality sleep	Avoid respiratory depressant substances
	OSA, EDS	Usual measures for OSA (see Chapter 4.14.3)
Neuromuscular disease	OSA, EDS	As above
Tourette syndrome	Disrupted sleep	Treatment of tic disorder should improve sleep
	Sleepwalking	
Cardiovascular disease		
Congestive heart failure	Central sleep apnoea, orthopnoea	Some antihypertensive, hypolipidaemic and antiarrythmic drugs can cause insomnia
Coronary heart disease	Nocturnal angina	Beta blockers can produce insomnia and nightmares
Respiratory disease		
COPD	Nocturnal dyspnoea	Hypnotics contraindicated in respiratory disease
Asthma	Nocturnal awakenings, EDS	Theophylline can cause insomnia
Gastrointestinal disorders (peptic ulcer, reflux oesophagitis)	Nocturnal awakenings	
Rheumatological disorders and other chronic pain conditions (including cancer)	Insomnia, disrupted sleep, EDS	Cortico-steroids and NSAI agents can disturb sleep Major analgesics have sedative effects Various anti-cancer drugs can disrupt sleep
Iron deficiency	PLMS	Treat underlying condition
	RLS	
Endocrine disease		
Diabetes	Awakenings from nocturnia or painful peripheral neuropathy	
	OSA	
Hyperthyroidism	Insomnia	
Myxoedema	OSA	
Chronic renal failure	Sleep disruption	Improvement with haemodialysis
	EDS	
	OSA	
	RLS and PLMS	
ICU patients	Severe sleep disruption and deprivation including sleep–wake cycle disorders causing confusional and psychotic states	Cause of reason for intensive care relevant, plus sleep disrupting effect of ICU environment, procedures and medications
Obesity	OSA	Complications of obesity (e.g. diabetes, joint diseases) affect sleep

KEY

AED = Anti-epileptic drugs
COPD = Chronic obstructive pulmonary disease
EDS = Excessive daytime sleepiness
ICU = Intensive care unit
NSAI = Non-steroidal anti-inflammatory
OSA = Obstructive sleep apnoea
PLMS = Periodic limb movements in sleep
RBD = REM sleep behaviour disorder
RLS = Restless legs syndrome

condition itself, although additional treatment for the sleep disorder may well be required.

- Treatment for other, co-morbid conditions may well be needed, in particular anxiety and depression which will have their own adverse effects on sleep.
- Hypnotic drugs should be used very sparingly (especially in the elderly) because of their potential complications as described in Chapter 6.2.2. In view of their respiratory-depressand effects, benzodiazepines in particular are contraindicated in sleep apnoea and in the presence of other severe respiratory disease.
- The possible effects on sleep of over-the-counter medications also need to be considered. Nasal decongestants and anorectics are stimulants and can cause insomnia. The same is true of caffeine-containing drinks used as 'energy boosters'. Non-prescribed sleeping preparations (which usually contain anti-histamines) may well cause daytime drowsiness.

As mentioned earlier, the link between medical illness and sleep disturbance may operate in the opposite direction. For example, chronic sleep loss or disruption is associated with the development of such chronic health problems as coronary artery disease or diabetes.

Further information

Kryger, M.H., Roth, T., and Dement, W.C. (eds.) (2005). *Principles and practice of sleep medicine* (4th edn). Elsevier Saunders, Philadelphia, PA.

American Academy of Sleep Medicine. (2005). *The international classification of sleep disorders: diagnostic and coding manual* (2nd edn). American Academy of Sleep Medicine, Westchester, IL.

Colton, H.R. and Altevogt, B.M. (eds.) (2006). *Sleep disorders and sleep deprivation: an unmet public health problem*. National Academies Press, Washington, DC. Available from http.//www.nap.edu/catalog/11617 html

Shapiro, C. and McCall Smith, A. (eds.) (1997). *Forensic aspects of sleep*. Wiley, Chichester.

References

1. Bonnet, M.H. (2005). Acute sleep deprivation. In *Principles and practice of sleep medicine* (4th edn) (eds. M.H. Kryger, T. Roth, and W.C. Dement), pp. 51–6. Saunders, Philadelphia, PA.
2. Dinges, D.F., Rogers, N.L., and Baynard, M.D. (2005). Chronic sleep deprivation. In *Principles and practice of sleep medicine* (4th edn) (eds. M.H. Kryger, T. Roth, and W.C. Dement), pp. 67–76. W.B. Saunders, Philadelphia, PA.
3. Stores, G. (2007). Clinical diagnosis and misdiagnosis of sleep disorders (Review). *Journal of Neurology, Neurosurgery, and Psychiatry*, **78**, 1293–7.
4. Buysse, D.J., Reynolds, C.F., Monk, T.H., *et al.* (1989) The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatric Research*, **28**, 193–213.

4.14.2 Insomnias

Colin A. Espie and Delwyn J. Bartlett

Introduction

Most people's experiences of poor sleep are memorable, because sleeplessness and its daytime consequences are unpleasant. There are those, however, for whom insomnia is the norm. Persistent and severe sleep disturbance affects at least one in 10 adults and one in five older adults, thus representing a considerable public health concern. Sleep disruption is central to a number of medical and psychiatric disorders, and insomnia is usually treated by general practitioners. Therefore differential diagnosis is important, and respiratory physicians, neurologists, psychiatrists, and clinical psychologists need to be involved. The purpose of this chapter is to summarize current understanding of the insomnias, their appraisal, and treatment. Particular emphasis will be placed upon evidence-based practical management.

Clinical features

Insomnia often remains unreported, and finally presents when a poor sleep pattern is well established. Alcohol has long been a first-line self-administered sleep aid, and recent years have seen an increasing use of 'over-the-counter' preparations and 'self-help' strategies. The clinical presentation is commonly of a frustrated patient, trapped in a vicious circle of anxiety and poor sleep, who reports having 'tried everything'.[1,2]

There may be concern about the pattern of sleep. This is the most quantifiable aspect of self-report relating to, for example, length of time taken to fall asleep, frequency and duration of wakenings, or total amount of sleep. A poorly established sleep pattern can lead to unpredictability of what sleep will be like on any given night. Patients often report poor quality of sleep, and subjective perceived quality can be a critically important outcome variable. Typical reports relate to light sleep and sleep felt to be unrestorative. Although it may be unclear how such complaints relate to EEG sleep architecture, the clinician should not overlook qualitative report as it may reflect underlying pathophysiology. Concerns are normally expressed also about the daytime effects of poor sleep. These can be cognitive effects, such as fatigue, sleepiness, inattention, and some impairments in performance, or emotional effects, such as irritability and anxiety.[2]

Classification

The *International Classification of Sleep Disorders* (second edition: **ICSD-2**)[3] was published in 2005 and provides the most comprehensive account of sleep disorders, both for descriptive purposes and for differential diagnosis (see Chapter 4.14.1). ICSD-2 describes insomnias as disorders of initiating and maintaining sleep. Patients may have either sleep-onset problems or wakenings from sleep, or both of these. Table 4.14.2.1 summarizes the principal classifications that relate to the insomnias, along with some other sleep disorders where patients commonly present with insomnia symptoms. As can be seen, concomitant symptomatology, potential aetiological factors, and sleeplessness require careful assessment in order to reach a valid diagnosis.

Table 4.14.2.1 The classification and differential diagnosis of the insomnias within ICSD-2

Classification	Sleep disorder	Essential features, complaint of insomnia plus
Insomnias	Psychophysiological insomnia	Learned sleep preventing associations, conditioned arousal, 'racing mind' phenomenon
	Paradoxical insomnia	Complaint of poor sleep disproportionate to sleep pattern and sleep duration
	Idiopathic insomnia	Insomnia typically begins in childhood or from birth
	Insomnia due to a mental disorder	Course of sleep disturbance concurrent with mental disorder
	Inadequate sleep hygiene	Daily living activities inconsistent with maintaining good-quality sleep
	Insomnia due to a medical disorder	Course of sleep disturbance concurrent with mental disorder
	Insomnia due to drug or substance	Sleep disruption caused by prescription medication, recreational drug, caffeine, alcohol or foodstuff
	Adjustment insomnia	Presence of identifiable stressor, insomnia resolves or is expected to resolve when stressor is removed
Sleep-related breathing disorders	Obstructive sleep apnoea syndrome	Excessive sleepiness, obstructed breathing in sleep, associated symptoms include snoring and a dry mouth
	Periodic limb movement disorder	Episodes of repetitive, highly stereotyped limb movements occurring in sleep
	Restless legs syndrome	Strong, nearly irresistible urge to move legs relieved by walking
Circadian rhythm sleep disorders	Delayed sleep phase type	Phase delay of major sleep episode, initial insomnia, excessive sleepiness in morning
	Advanced sleep-phase type	Phase advance of major sleep episode, inability to stay awake in evening, early wakening

Diagnosis and differential diagnosis

Severity of insomnia is judged along dimensions of frequency, intensity, and duration, as well as impact on daytime functioning and quality of life. Generally, the criteria for severe and chronic insomnia are a minimum duration of 6 months with problems presenting three or more nights per week. Restlessness, irritability, anxiety, daytime fatigue, and tiredness commonly accompany such presentations.[2] Mild and moderate insomnia may be diagnosed where problems are less intrusive.

Most patients presenting with insomnia have psychophysiological difficulty initiating and/or maintaining sleep. Usually marked functional effects and somatized tension associated with sleep are evident. The patient reports extreme tiredness while being unable to sleep satisfactorily and appears preoccupied with sleep and its consequences. This contrasts, for example, with the circadian disorders where, in delayed sleep-phase type, the patient may not feel sleepy until late in the normal sleep period, and in advanced sleep-phase type, may waken early and be unable to return to sleep. Taking a history, incorporating screening questions on restlessness, limb movements, and breathing can help to diagnose obstructive sleep apnoea syndrome, periodic limb movement disorder, and restless legs syndrome, although full polysomnographic evaluation may also be required.[4] However, polysomnography is not essential for the diagnosis of insomnia, for which sleep diary monitoring (see Chapter 4.14.1) is usually the most useful form of assessment.[2] Wrist actigraphy is an inexpensive objective evaluation, which estimates sleep/wakefulness based upon body movement.[5] Continuous recordings can be made over 5 to 10 consecutive 24 h periods. It is useful in identifying paradoxical insomnia, and charted data can be inspected for circadian anomalies.

Other causes of insomnia are reported in Table 4.14.2.1 and should not be overlooked. In particular, insomnias due to a drug or substance can include hypnotic-dependent sleep disorder, associated most commonly with benzodiazepine (BZ) drugs where withdrawal leads to exacerbation of the primary problem.[6] This can be mistaken for a severe underlying insomnia and hence reinforce hypnotic dependency. Likewise, a wide range of psychiatric conditions, particularly affective disorders, has associated sleep symptomatology (see Chapter 4.14.1). A primary diagnosis of psychophysiological insomnia cannot be made where diagnostic criteria for DSM-IV Axis I or Axis II disorders are fulfilled. However, it is very important to note that sleep disturbance often precedes depression. The bulk of the psychiatric epidemiological data indicate that insomnia is an independent risk factor for first episode depressive illness, and for recurrence of depression, in adults of all ages.[7,8] Insomnia should not be assumed to be simply a symptom of underlying depression, even when depression is present. Unless the illness courses clearly co-vary it is best to make a diagnosis of co-morbid insomnia. Similar caveats apply to insomnia associated with medical disorders, both in terms of identifying a primary illness, and concluding that insomnia has *the* status of an associated/ co-morbid disorder (see Chapter 4.14.1).

Epidemiology

Insomnia affects one-third of adults occasionally, and 9 to 12 per cent on a chronic basis. It is more common in women, in shift workers, and in patients with medical and psychiatric disorders. Prevalence amongst older adults has been estimated at up to 25 per cent and sleepiness and hypnotic drugs are risk factors for injury and fracture.[9] The decline in prescription of anxiolytics has been greater than the rate of decline for hypnotics [taking BZ and benzodiazepine receptor agonists (BzRAs) together]. Furthermore, there is increasing use of (off-label) sedative antidepressants primarily to treat insomnia.

Aetiology

Many patients report having been marginal light sleepers before developing insomnia.[4] Sleep disturbance often arises during life change or stress, and such adjustment sleep disorder may represent a normal transient disruption of sleep. However, secondary factors, such as anxiety over sleep and faulty sleep–wake conditioning, may

exacerbate and maintain the insomnia as a chronic problem when sleep itself becomes a focus for concern. People with insomnia may be hyperaroused relative to normal sleepers, for example having higher levels of cortisol and ACTH, and also find it difficult to 'down-regulate' their arousal at bedtime.[2,10,11]

Course and prognosis

There has been little research on the natural course of insomnia. However, untreated psychophysiological insomnia can last for decades, and may gradually worsen over time. Indeed, there is a developmental trend for sleep pattern to deteriorate, with increasing age. On the other hand, delayed sleep-phase syndrome and insufficient sleep hygiene can be associated with lifestyle problems and may ameliorate as these are resolved. Although certain insomnias *tend* to persist if untreated, prognosis with effective treatment can be very good.

Treatment

A review of the evidence

(a) Drug therapy

Traditionally, insomnia has been treated pharmacologically. Barbiturates were superseded by BZ compounds during the 1960s and 1970s. These drugs were safer in overdose, were thought to have fewer side effects, and to be less *addictive*. Controlled studies have demonstrated that a considerable number of BZ, of short to intermediate half-life, are effective hypnotic agents. However, from the mid-1970s potential problems became apparent, both during administration and withdrawal. Longer-acting hypnotics were prone to carry-over effects of morning lethargy, and shorter-acting drugs to 'rebound insomnia'.[6] Furthermore, tolerance develops, leading either to increased dosing or switching to alternative medication. Although BZs used for short periods/intermittently can maintain effectiveness, these are not the treatment of choice in chronic insomnia,[12] and are contraindicated in older adults and where insomnia may involve sleep-related breathing disorder because of their potentially depressant effects on respiration. A number of BZ compounds have been removed from the market in the United Kingdom, United States, and elsewhere.

Contemporary hypnotic therapy has extended to include BzRAs (often referred to as the 'z' drugs), and more recently melatonin receptor agonists (MeRAs) have been introduced. Whereas the place in therapeutics of MeRAs has yet to become established, the BzRAs are often thought to offer more sustained benefit for insomnia, and to have fewer adverse effects. Nevertheless, there remains uncertainty about the effectiveness of BzRAs in chronic insomnia.[13]

(b) Psychological therapy

Psychological treatment for chronic insomnia, primarily in the form of cognitive behavioural therapy (CBT), has been extensively investigated in over 100 controlled studies during the past 20 years. Five meta-analyses and numerous systematic reviews have demonstrated that CBT is associated with large effect size changes (measured in standardized z scores) in the primary symptom measures of sleep latency (difficulty getting to sleep) and wake time after sleep-onset (difficulty remaining asleep).[14,15] Around 70 per cent of patients with persistent sleep problems appear to benefit from CBT and effects are maintained to long-term follow-up. It is thought that CBT achieves these outcomes because it tackles directly the dysfunctional thoughts and maladaptive behaviours that otherwise maintain insomnia. Recent controlled studies have shown that CBT may be effective in general practice settings with nurses delivering the intervention according to a standard protocol.[16,17] Despite the superior efficacy of CBT relative to medication for insomnia, and these recent demonstrations of CBT working in real-world settings, practical problems remain in making CBT widely available.

Within the CBT model, a number of strategies have strong empirical support. Behavioural procedures such as stimulus control and sleep restriction, and cognitive strategies such as paradoxical intention and thought restructuring have been extensively investigated[2,14,15] and are outlined briefly below.

(c) Melatonin, light therapy, and exercise

The pineal hormone melatonin has been the subject of highly publicized claims. However, scientific research has been limited. Several controlled studies support its sleep-promoting effects, but the use of melatonin continues to be controversial. At best it may be useful as a chronobiotic for reducing sleep latency.[18] Several MeRA products are currently under formal evaluation, so more data may be available soon.

Bright light is a potent marker for human circadian rhythms, and has been known for some time to enable the resetting of such rhythms in advanced sleep-phase syndrome and delayed sleep-phase syndrome.[19] The results of studies investigating the efficacy of bright light against psychological treatments for psychophysiological insomnia are awaited. A limiting factor to the value of light therapy is that continued treatment may be required to maintain therapeutic effects.

Athletic people sleep well, although this may be more to do with behavioural patterning than aerobic fitness. Nevertheless, there is evidence that exercise can have positive effects upon sleep quality, particularly if taken late afternoon or early evening, and in otherwise relatively fit individuals.[20] Morning exercise can also be an effective modality to encourage the same waking time and early morning light exposure; which help to reset sleep patterns on a daily basis.

Advice about management

(a) General perspective

Non-pharmacological treatment using CBT procedures should be preferred over pharmacological treatment, in cases of severe persistent insomnia. Hypnotic agents should be recommended mainly for short-term or occasional use, although longer-term trial data are now becoming available. The practitioner should be aware of morning-after effects, and potential problems of withdrawal and dependency, not only with BZs but also possibly with BzRAs. Psychological intervention may also facilitate reduction or discontinuation of medication in hypnotic-dependent person with insomnias.[21] There is limited support for the use of melatonin or exercise as treatments of choice, although light therapy seems effective for circadian disorders.

(b) Using cognitive behavioural therapies

Brief descriptions of effective management strategies are presented in Tables 4.14.2.2 and 4.14.2.3. The following text provides

Table 4.14.2.2 Summary description of sleep hygiene and education components for the treatment of chronic insomnia

Components of sleep education
The need for sleep and its functions
Sleep patterns across the lifespan
Sleep as a process with stages/phases
Factors adversely affecting steep
The effects of sleep loss
The concept of insomnia
Measuring sleep and sleep problems
Components of sleep hygiene treatment
Bedroom comfortable for sleep
Regular exercise, timing, and fitness
Stable and appropriate diet
Undesirable effects of caffeine and other stimulants
Moderation of alcohol consumption
Other common 'self-help' strategies

explanation of underlying psychological models and further information on implementation.

(i) Sleep education and sleep hygiene

The simple provision of information ameliorates the sense of being out of control. Inaccurate attributions are challenged and misunderstandings corrected by understanding what sleep is, how ommon insomnia can be, how sleep changes with age, good sleep hygiene practices, and some facts about sleep loss. Similarly, sleep hygiene provides patients with a starting point for self-management. These techniques are best construed as introductory but they will not of themselves treat insomnia effectively.

Table 4.14.2.3 Summary description of cognitive behavioural components for the treatment of chronic insomnia

Components of stimulus control and sleep restriction treatment
Define individual sleep requirements
Establish parameters for bedtime period (threshold time and rising time)
Eliminate daytime napping
Differentiate rest from sleep
Schedule sleep periods with respect to needs
Establish 7 day per week compliance
Remove incompatible activity from bedroom environment
Rise from bed if wakeful (>20 min)
Avoid recovery sleep as 'compensation'
Establish stability from night to night
Adjust the sleep period as sleep efficiency improves
Components of cognitive intervention
Identify thought patterns and content that intrude
Address (mis)attributions connecting sleep and waking life
Establish rehearsal/planning time in early evening
Relaxation and imagery training
Distraction and thought blocking
Develop accurate beliefs/attributions about sleep and sleep loss
Challenge negative and invalid thoughts
Eliminate 'effort' to control sleep
Motivate to maintain behaviour and cognitive change
Utilize relapse-prevention techniques

(ii) Stimulus control treatment

Stimulus control increases the bedroom's cueing potential for sleep. For good sleepers, the pre-bedtime period and the stimulus environment trigger positive associations of sleepiness and sleep. For the poor sleeper, however, the bedroom triggers associations with restlessness and lengthy night-time wakening via a stimulus-response relationship, thereby continuing to promote wakefulness and arousal. The model is similar to phobic conditions where a conditioned stimulus precipitates an anxiety response.

Treatment involves removing from the bedroom all stimuli which are potentially sleep-incompatible. Reading and watching television, for example, are confined to living rooms. Sleeping is excluded from living areas and from daytime, and wakefulness is excluded from the bedroom. The individual is instructed to get up if not asleep within 15–20 min or if wakeful during the night. Conceptually, stimulus control is a reconditioning treatment which forces discrimination between daytime and sleeping environments.

(iii) Sleep restriction therapy

Sleep restriction restricts sleep to the length of time which the person is likely to sleep. This may be equivalent to promoting 'core sleep' at the expense of 'optional sleep'. Sleep restriction primarily aims to improve sleep efficiency. Since sleep efficiency is the ratio of time asleep to time in bed, it can be improved either by increasing the numerator (time spent asleep) or by reducing the denominator (time spent in bed). People with insomnia generally seek the former, but this may not be necessary, either biologically or psychologically. Sleep restriction first involves recording in a sleep diary and calculating average nightly sleep duration. The aim, then, is to obtain this average each night. This is achieved by setting rising time as an 'anchor' each day and delaying going to bed until a 'threshold time' which permits this designated amount of sleep. Thus, the sleep period is compressed and sleep efficiency is likely to increase. The permitted 'sleep window' can then be titrated week-by-week in 15 increments in response to sleep efficiency improvements.

(iv) Cognitive control

This technique aims to deal with thought material in advance of bedtime and to reduce intrusive bedtime thinking. The person with insomnia is asked to set aside 15 to 20 min in the early evening to rehearse the day and to plan ahead for tomorrow; thus putting the day to rest. It is a technique for dealing with unfinished business and may be most effective for rehearsal, planning, and self-evaluative thoughts which are important to the individual and which, if not dealt with, may intrude during the sleep-onset period.

(v) Thought suppression

Thought-stopping and articulatory suppression attempt to interrupt the flow of thoughts. No attempt is made to deal with thought material *per se* but rather to attenuate thinking. With articulatory suppression the patient is instructed to repeat, subvocally, the word 'the' every 3 s. This procedure is derived from the experimental psychology literature. Articulatory suppression is thought to occupy the short-term memory store used in the processing of

information. The type of material most likely to respond is repetitive but non-affect-laden thoughts, not powerful enough to demand attention. Additionally, this technique may be useful during the night to enable rapid return to sleep.

(vi) Imagery and relaxation

There is a wide range of relaxation methods including progressive relaxation, imagery training, biofeedback, meditation, hypnosis, and autogenic training, but little evidence to indicate superiority of any one approach. Furthermore, there is little evidence to support either the presumption that people with insomnia are hyperaroused in physiological terms, or that relaxation has its effect through autonomic change. At the cognitive level, these techniques may act through distraction and the promotion of mastery. During relaxation, the mind focuses upon alternative themes such as visualized images or physiological responses. In meditation the focus is upon a 'mantra' and in self-hypnosis upon positive self-statements. Relaxation may be effective for thought processes that are anxiety-based, confused, and which flit from topic-to-topic.

(vii) Cognitive restructuring

Cognitive restructuring challenges faulty beliefs which maintain wakefulness and the helplessness which many people with insomnia report. It appears to work through appraisal by testing the validity of assumptions against evidence and real-life experience. As an evaluative technique, it may be effective with beliefs that are irrational but compelling. If such thoughts, for example 'I am going to be incapable at work tomorrow', are not challenged, they will create high levels of preoccupation and anxiety and sleep is unlikely to occur. With cognitive restructuring, the person with insomnia learns alternative responses to replace inaccurate thinking.

(viii) Paradoxical intention

Finally, the technique of paradoxical intention is useful in situations where performance anxiety has developed, that is, where the effort to produce a response inhibits that response itself. The paradoxical instruction is to allow sleep to occur naturally through passively attempting to remain quietly wakeful rather than attempting to fall asleep. Paradox may be regarded as a decatastrophizing technique since it appears to act upon the ultimate anxious thought (of remaining awake indefinitely) initially by focusing on and enhancing this thought (a habituation model) and then subjecting it to appraisal through rationalization and experience. By intending to remain awake, and failing to do so, the strength of the sleep drive is re-established, and performance effort is reduced.

Possibilities for prevention

There is insufficient knowledge of the natural course of transient sleep disorders. Mention has been made of adjustment sleep disorder and of the association of life events and stressors with the onset of insomnia. Systematic research is required to establish the 'setting conditions' for the secondary maintenance of insomnia beyond an initial normative reaction to events. Perhaps there is an interaction with a predisposing tendency to light sleep, or with introspection and worry. The instinct to increase opportunity to sleep (spend longer in bed to catch up) when insomnia symptoms develop should probably be resisted. If anything it may be better to advise patients to limit sleep opportunity so that their pattern knits together again more quickly.

The establishment and maintenance of a regular 'tight' routine, both pre-bedtime and in terms of sleep schedule, seem to be important preventive factors. Such chronobehavioural functioning can be at risk of disruption by, for example, jet lag, shift work, weekend patterns differing from weekday, adolescent lifestyle, and retirement. Adherence to, and/or reinstatement of, an adaptive pattern seems crucial.

It is important not to underestimate the importance of attitudes and beliefs in the presentation of insomnia. Exaggerated or emotionally and mentally arousing thoughts should be dealt with promptly. Sleep loss can be distressing, but patients should be reminded that nature seeks to restore equilibrium. What they need to do is to provide the conditions under which sleep can occur rather than attempt directly to control the sleep process. Expectations are important also, since it is the breach of these which generally give rise to anxiety and dysfunctional beliefs about sleep requirements. More often than not sleep-related expectations are unrealistic and require reappraisal, even more so in older adults.

Finally, prevention should be extended to the known extrinsic causes of certain sleep disorders. Where alcohol, stimulants, or proprietary drugs interfere with sleep and the recovery of the normal sleep process, attention should be paid to these factors. Better still, patients should be encouraged to seek advice early rather than go down the path of self-administered treatment. Avoiding the use of hypnotic agents, both in general practice and during acute admissions to hospital, would substantially reduce the number of iatrogenic cases of chronic insomnia.

Further information

Espie, C.A. (2006). *Overcoming insomnia and sleep problems: a self-help guide using cognitive behavioral techniques*. Constable & Robinson Ltd, London [ISBN13: 978-184529-070-2/ ISBN10: 1-84529-070-4].

Perlis, M.L. and Lichstein, K.L. (2003). *Treating sleep disorders: principles and practice of behavioral sleep medicine*. John Wiley & Sons, Inc. Hoboken, NJ [ISBN0-47-44343-3].

Morin, C.M., Bootzin, R.R., Buysse, D.J., *et al.* (2006). Psychological and behavioural treatment of insomnia. Update of the recent evidence (1998–2004) prepared by a Task Force of the American Academy of Sleep Medicine. *Sleep*, **29**, 1398–414.

Morin, C.M. and Espie, C.A. (2003). *Insomnia: a clinical guide to assessment and treatment*. Kluwer Academic/ Plenum Publishers, New York [ISBN 0-306-47750-5].

References

1. Espie, C.A. (1991). *The psychological treatment of insomnia*. Wiley, Chichester.
2. Morin, C.M. and Espie, C.A. (2003). *Insomnia: a clinical guide to assessment and treatment*. Kluwer Academic/Plenum Publishers, New York.
3. American Academy of Sleep Medicine. (2005). *International classification of sleep disorders: diagnostic and coding manual* (2nd edn). AASM, Westchester, IL.
4. Reite, M., Buysse, D., Reynolds, C., *et al.* (1995). The use of polysomnography in the evaluation of insomnia. *Sleep*, **18**, 58–70.
5. Ancoli-Israel, S., Cole, R., Alessi, C., *et al.* (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, **26**, 342–92.
6. Kripke, D. (2000). Hypnotic drugs: deadly risks, doubtful benefits. *Sleep Medicine Reviews*, 2000, **4**, 5–20.

7. Cole, M.G. and Dendukuri, N. (2003). Risk factors for depression among elderly community subjects: a systematic review and meta-analysis. *The American Journal of Psychiatry*, **160**, 1147–56.
8. Riemann, D. and Voderholzer, U. (2003). Primary insomnia: a risk factor to develop depression? *Journal of Affective Disorders*, **76**, 255–9.
9. Lichstein, K.L., Durrence, H.H., Reidel, B.W., *et al.* (2004). *The epidemiology of sleep: age, gender and ethnicity*. Lawrence Erlbaum Associates, Mahwah, NJ.
10. Espie, C.A. (2002). Insomnia: conceptual issues in the development, persistence and treatment of sleep disorder in adults. *Annual Review of Psychology*, **53**, 215–43.
11. Perlis, M.L., Pigeon, W., and Smith, M.T. (2005). Etiology and pathophysiology of insomnia. In *The principles and practice of sleep medicine* (4th edn) (eds. M.H. Kryger, T. Roth, and W.C. Dement), pp. 714–25. W.B. Saunders, Philadelphia.
12. NIH. (2005). *State-of-the-science conference statement on manifestations and management of chronic insomnia in adults*. National Institutes of Health, Vol. 22, Number 2. Natcher Conference Center, Bethesda Maryland, USA.
13. NICE. (2004). *Guidance on the use of zaleplon, zolpidem and zopiclone for the short-term management of insomnia*. Technology Appraisal Guidance No.77. National Institute for Clinical Excellence, London.
14. Smith, M.T., Perlis, M.L., Park, A., *et al.* (2002). Behavioral treatment vs pharmacotherapy for insomnia—a comparative meta-analysis. *The American Journal of Psychiatry*, **159**, 5–11.
15. Morin, C.M., Bootzin, R.R., Buysse, D.J., *et al.* (2006). Psychological and behavioural treatment of insomnia. Update of the recent evidence (1998–2004) prepared by a Task Force of the American Academy of Sleep Medicine. *Sleep*, **29**, 1398–414.
16. Espie, C.A., Inglis, S.J., Tessier, S., *et al.* (2001). The clinical effectiveness of cognitive behaviour therapy for chronic insomnia: implementation and evaluation of a Sleep Clinic in general medical practice. *Behaviour Research and Therapy*, **39**, 45–60.
17. Espie, C.A., MacMahon, K.M.A., and Kelly, H.L. (2007). Randomised clinical effectiveness trial of nurse-administered small group CBT for persistent insomnia in general practice. *Sleep*, **30**, 574–84.
18. Mendelson, W.B. (1997). A critical evaluation of the hypnotic efficacy of melatonin. *Sleep*, **20**, 916–19.
19. Czeisler, C.A., Johnson, M.P., Duffy, J.F., *et al.* (1990). Exposure to bright light and darkness to treat physiologic maladaptation to night work. *The New England Journal of Medicine*, **322**, 1253–8.
20. Singh, N.A., Clements, K.M., and Fiatarone, M.A. (1997). A randomized controlled trial of the effect of exercise on sleep. *Sleep*, **20**, 95–101.
21. Morgan, K., Dixon, S., Mathers, M., *et al.* (2003). Psychological treatment for insomnia in the management of long-term hypnotic drug use: a pragmatic randomised controlled trial. *The British Journal of General Practice*, **53**, 923–8.

4.14.3 Excessive sleepiness

Michel Billiard

Introduction

Excessive sleepiness is not an homogeneous concept. It can manifest itself as bouts of sleepiness, irresistible and refreshing sleep episodes, abnormal lengthening of night sleep with a major difficulty waking up in the morning or at the end of a nap or even periods of a week or so of almost continuous sleep recurring at several months' intervals.

According to the recent second edition of the *International Classification of Sleep Disorders* (*ICSD*-2),[1] disorders of excessive sleepiness are distributed within three chapters: sleep-related breathing disorders, hypersomnias of central origin not due to a circadian rhythm sleep disorder, sleep-related breathing disorders, or other cause of disturbed nocturnal sleep, and circadian rhythm sleep disorders.

However in this volume aimed at psychiatrists, the presentation of disorders of excessive sleepiness will obey another logic. Following "Generalities" including epidemiology, morbidity, clinical work-up, and laboratory tests, the various aetiologies will be presented according to the following six subchapters:

- Hypersomnia not due to substance or known physiological condition (non-organic hypersomnia or psychiatric hypersomnia)
- Hypersomnia due to drug or substance
- Behaviourally induced insufficient sleep syndrome
- Hypersomnia in the context of sleep-related breathing disorders
- Hypersomnias of central origin
- And the special case of delayed sleep phase syndrome.

Epidemiology

Contrary to common thinking, excessive sleepiness is neither exceptional nor rare. Epidemiological surveys generally agree on a figure of severe sleepiness (daily and embarrassing) in 5 per cent of the general population and of moderate sleepiness (occasional) in another 15 per cent.[2] Interestingly, only a fraction of these subjects are aware of their condition, due to the fact that they progressively lose reference to a normal state of alertness. As a consequence many subjects will not consult their physician for excessive sleepiness but will be brought to him by the spouse, worried about his or her falling asleep repeatedly in the middle of the day, or even referred by the company's doctor due to unexplained car accidents or poor work efficiency.

Morbidity

Excessive sleepiness has a severe impact on the life of patients. Nearly half of the patients with excessive sleepiness report automobile accidents. Many have lost jobs because of their sleepiness. In addition, sleepiness is disruptive of family life. Cognitive function is also impaired by sleepiness. In children excessive sleepiness has been associated with learning disability and in adults memory problems are frequent.

Clinical work-up and laboratory tests

Whatever the circumstance of the first visit, the patient should be interviewed on the history of excessive sleepiness, the type and severity of it, the associated symptoms, the familial and occupational consequences, the past and current treatments, and the personal and familial medical past-history.

In addition, the subject will complete a self-administered behavioural scale, the Epworth sleepiness scale. This scale asks the subject to rate the probability of dozing from 0 (would never doze) to

3 (high chance of dozing) in eight more or less soporific daily situations. A score of over 10 is taken to indicate abnormal sleepiness.

The subject will then undergo a physical and psychological examination.

Laboratory tests will be chosen according to the clinical impression.

The most frequently used test is the *multiple sleep latency test* (MSLT). The test was developed on the basis of the following principle. The sleepier the subject, the faster he falls asleep. The test is based on 20 min polygraphic recordings (EEG, EOG, EMG) repeated every 2 h (four or five times a day) starting 2 h about after morning awakening. The global sleepiness index is provided by the mean latency to sleep in the four or five tests. A sleep laboratory of less than 5 min indicates pathologic sleepiness, a sleep latency from 10 to 20 min is considerd as normal, and latencies falling between the normal and the pathological values are considered as a diagnostic grey area.

Another test, the *maintenance of wakefulness test* (MWT), is a variant of the MSLT. It was designed to evaluate treatment efficiency in patients with excessive sleepiness. The major difference with the MSLT is in the instruction given to the test subject. The subject being tested is told to attempt to remain awake. The subject is seated in comfortable position in bed, as opposed to lying down in the MSLT, with low lighting behind him (7.5 W at 1 m). Specific recommendations include using a four-trial, 40 min version of the MWT. A mean sleep latency of less than 8 min on the 40 min MWT is abnormal and scores between 8 and 40 min are of uncertain significance.

Prolonged polysomnographic recordings, obtained by either traditional laboratory polysomnographic monitoring or ambulatory recordings, provide a good picture of the actual time asleep within the 24 h period. However, this procedure is neither validated nor standardized.

In addition, whenever there is some doubt about the possibility of hypersomnia associated with a psychiatric disorder, a *psychometric/psychiatric evaluation* will be performed.

Aetiology and treatment

Hypersomnia not due to substance or known physiological condition

It explains about 5 to 7 per cent of cases of hypersomnia seen in sleep disorders centres. Women are more susceptible than men.

Excessive daytime sleepiness is reported. Subjects show an elevated score on the Epworth sleepiness scale. Night sleep is perceived as non-restorative and generally of poor quality. Patients are often intensely focused on their hypersomnia, and psychiatric symptoms typically become apparent only after prolonged interview or psychometric testing. Poor work attendance, abruptly leaving work because of a perceived need to sleep are common. Polysomnography typically shows a prolonged sleep latency, an increased wake time after sleep onset, and a low sleep efficiency. REM latency may be shortened in the case of bipolar disorder. Contrasting with the elevated score on the Epworth sleepiness scale, sleep latency on the MSLT is often within normal limits. A 24 h continuous sleep recording typically shows considerable time spent in bed during day and night, a behaviour referred to as clinophilia, from the Greek κλινη (bed) and φιλεω (love).

Psychiatric interview is essential to diagnose the underlying condition. Causative psychiatric conditions include bipolar type II disorder, dysthymic disorder, undifferentiated somatoform disorder, adjustment disorder, or personality disorder.

Conventional drugs such as antidepressants or anxiolytics are often insufficient. Modafinil, an awakening drug given at a daily dose of 100 to 200 mg, is usually active.

In the group of psychiatric disorders a separate place should be reserved to seasonal affective disorder remarkable for episodes of major depression occurring only during the winter months, associated with fatigue, loss of concentration, increased appetite for carbohydrates, weight gain, and increased sleep duration. Morning bright light treatment (2500 lux for 2 h) is efficient.

Hypersomnia due to drug or substance

A wide spectrum of medications used in psychiatry may be responsible for excessive sleepiness.

(a) Anxiolytics and hypnotics

Benzodiazepines have sedative effects, but these effects vary with dose, administration (single or repeated dose), age, and state of the subject (normal, anxious, or depressed). Non-benzodiazepines usually induce limited sleepiness only.

(b) Antidepressants

Tricyclic antidepressants have sedative properties depending on the molecule, dose, and the subject to whom they are administered. SSRI can also induce sleepiness with high within-patient variability. Venlafaxin, a serotonine, and norepinephrine reuptake blocker may induce excessive sleepiness.

(c) Neuroleptics

The degree of sedation varies widely from subject to subject. Empirically, three-fourth of the patients treated with neuroleptic phenothiazines experience sleepiness in a dependent manner. Among the newer agents clozapine is the most sedating drug, followed by olanzapine and quetiapine. Risperidone and sertindole are less sedating drugs.

Behaviourally induced insufficient sleep syndrome

According to ICSD-2, this syndrome occurs when an individual persistently fails to obtain the amount of sleep required to maintain normal levels of alertness. Behaviourally induced insufficient sleep syndrome is likely the most common cause of daytime sleepiness. In a population-based study conducted in Japan among 3030 subjects aged 20 years and older, 29 per cent slept less than 6 h per night, and 23 per cent reported having insufficient sleep.[3] The syndrome is likely to be widespread in truck drivers, working mothers, family doctors, executives, and students. The main symptoms are excessive sleepiness in the afternoon or early evening, decrease of diurnal performances, and, of interest to the psychiatrist, irritability, nervousness, and depression. Diagnosis of the syndrome is relatively easy provided that a thorough interview is conducted. The most rational treatment is an increase of daily total sleep time, either by spending more time in bed at night, or by taking one or two naps per day.

Hypersomnia in the context of sleep-related breathing disorders

The most frequent condition among these disorders is the obstructive sleep apnoea syndrome. This syndrome was first described by

Guilleminault *et al.*[4] It is most frequent in 50-year-old males. According to Young *et al.*[5] the prevalence of obstructive sleep apnoeas accompanied by excessive daytime sleepiness in North America is 4 per cent in men and 2 per cent in women.

Clinical features include night-time and daytime symptoms. Night-time symptoms are represented by loud snoring, apnoeic episodes ending with sonorous breathing resumption, nocturia, severe fatigue upon awakening, and sometimes headache. Daytime symptoms are dominated by excessive sleepiness, which varies in intensity among patients. Other symptoms include irritability, negligence, loss of concentration, loss of libido, impotence, and sometimes depression.

Patients are often obese or mildly obese. High blood pressure is a frequent feature. The ear, nose, and throat examination usually reveals a narrow upper airway due to close-set posterior tonsillar pillars, an abnormally long and hypotonic soft palate, a hypertrophic uvula, or macroglossia.

The positive diagnosis rests on polysomnography allowing the observation of nocturnal disrupted sleep and the identification of apnoeas and of their type (obstructive, central, or mixed) as well as their consequences on heart rate, oxygen desaturation, and degree of somnolence.

Of note, some subjects do not have apnoeas or hypopnoeas but increasing respiratory effort resulting in respiratory effort-related arousals (RERAs) and are believed to be as much at risk for complications. This state is most accurately identified with a quantative measurement of airflow and oesophageal manometry.

Obstructive sleep apnoea patients are at risk for systemic hypertension, occasional arrhythmias and conduction disturbances, cardiac or cerebral ischaemia, functional cognitive impairment, and depression. In a multicentre telephone survey carried-out in 1994–1999 in five European countries, among 18 980 subjects aged 15 to 100 years, 18 per cent of the individuals receiving a diagnosis of major depression also had a sleep-related breathing disorder and 17.6 per cent of the individuals receiving a diagnosis of sleep-related breathing disorder also had a diagnosis of major depression.[6]

There are several possible approaches to treatment:

- Weight loss may be beneficial. Avoidance of alcohol and sedatives should be recommended in all cases.
- Continuous positive airway pressure (CPAP) at night is the most widely used treatment. Good compliance requires proper preparation for the patient and an adaptation period.
- Oral appliances are suitable in mild obstructive sleep apnoea syndrome.
- Surgery consists mainly in nasal reconstruction in case of symptomatic airway blockade caused by bony, cartilaginous, or hypertrophied tissues that interfere with nasal breathing during sleep.

Hypersomnias of central origin

(a) Narcolepsy

First described in 1877[7] and given its name by Gelineau in 1880,[8] narcolepsy is now distinguished into three entities, narcolepsy with cataplexy, narcolepsy without cataplexy, and narcolepsy due to medical condition.

(i) Narcolepsy with cataplexy

Narcolepsy with cataplexy is not an exceptional condition. According to most recent evaluations its prevalence is 0.20 to 0.40 per 1000, i.e. slightly less than the prevalence of multiple sleep sclerosis.

Narcolepsy with cataplexy is characterized by two cardinal symptoms, excessive daytime sleepiness/irresistible sleep episodes and cataplexy, and other clinical features that are not necessarily part of the actual clinical picture.

Excessive sleepiness occurs daily. It comes in waves of varying degree of severity, depending on the individual, building up into irresistible and refreshing episodes of sleep. Excessive sleepiness is brought on by passive situations such as watching TV, being a passenger in a car, or attending a lecture. However it can also awkwardly occur in unexpected situations such as eating, walking, swimming, or driving a car. Excessive sleepiness may lead to automatic behaviour, such as saying totally inappropriate words or sentences, arranging objects in unlikely places, or driving a vehicle to an unintended destination.

Cataplexy is the most specific symptom. It consists of a sudden bilateral loss of voluntary muscle tone with preserved consciousness. It is triggered by environmental factors which are usually positive, such as a fit of laughter, receiving a compliment, humour expressed by the subject himself, the sight of prey for the hunter, the perception of a fish biting at the hook for the angler, a well-caught ball at tennis, or by anger, but almost never by stress or fear. All the striated muscles may be affected except the extra-ocular and respiratory musculature, leading to the progressive slackening of the whole body. More often however the attack is partial, involving certain muscles only, for example the jaw muscles, producing sudden difficulty in articulating words; the facial muscles, responsible for a grimace; or the thigh muscles, causing a brief unlocking of the knees. Consciousness is maintained during the episode. Cataplexy varies in duration from a split second to several minutes. Attack frequency varies from only a few per year, or even less, to several per day. In rare cases, especially after abrupt withdrawal of antidepressant medication, episodes of cataplexy may repeat themselves for minutes or hours, a state referred to as "status cataplecticus".

The other clinical features are deemed accessory to the extent that they are not indispensable for diagnosis.

Hallucinations, whether hypnagogic (at the onset of sleep) or hypnopompic (on awakening), are vivid perceptual experiences, not dreams, either visual, or auditory or kinetic. Patients may perceive someone entering and walking in the room, find themselves flying through the air or falling from a skyscraper. In some cases of unrecognized narcolepsy with daytime hypnagogic/hypnopompic hallucinations, the patient may be mistakenly diagnosed as having a delusional psychosis.

Sleep paralysis is a sudden inability to move during the transition from sleep to wakefulness or vice versa, while the subject is conscious. It is often accompanied by hypnagogic hallucinations. It is very unpleasant. In narcoleptic patients sleep paralysis may last up to 10 min.

Nocturnal sleep disruption occurs in approximately 50 per cent of cases, depending on age and the time elapsed since the onset of condition. The patient typically falls asleep as soon as he gets to bed, but his sleep is disturbed by recurring awakenings. He may complain of disturbing dreams.

REM-sleep behaviour disorder is frequent, either as a mere polysomnographic finding, excess of muscle tone or phasic EMG twitching activity during REM sleep, or as clinically significant complaint manifesting itself as an attempted enactment of distinctly altered, unpleasant, action-filled, and violent dreams in which the individual is being confronted, attacked, or chased by unfamiliar people. Periodic limb movements in sleep are more frequent than in normal subjects.

Physical examination is normal except for a frequently increased body mass index, especially in children at the onset of the condition. Noteworthy is the abolition of deep tendon reflexes during a cataplectic attack.

In a majority of cases, excessive daytime sleepiness is the first symptom to appear. In some patients, cataplexy occurs at the same time as excessive daytime sleepiness, but more often is delayed by one or several years. It is extremely rare for cataplexy to be the first clinical manifestation of narcolepsy with cataplexy.

The clinical diagnosis is based on clinical features. However, additional tests are highly recommended to confirm the diagnosis. The first test is polysomnography followed by an MSLT. At night a sleep onset REM period (SOREMP) is highly specific. It is observed in 25 to 50 per cent of cases. An increase in the amount of stage 1 NREM sleep and repeated awakenings are frequent findings. The MSLT shows a mean sleep latency of less than 8 min and two or more SOREMPs. However, some patients, especially in middle or old age, with clear-cut excessive daytime sleepiness and cataplexy, may have only one SOREMP, or even none during the MSLT procedure.

Today a CSF level of hypocretin-1 below 110 pg/mL is recognized as highly specific and sensitive for narcolepsy with cataplexy. However, up to 10 per cent of narcolepsy with cataplexy patients have normal levels of hypocretin-1.

There is no set pattern of evolution across patients. Excessive daytime sleepiness and irresistible episodes of sleep persist throughout life, but may diminish with age in some subjects. Cataplexy may spontaneously diminish with time or even totally disappear in some patients.

Narcolepsy is a very incapacitating disease. It interferes with every aspect of life. Education, performance at school and workplace, driving capability, recreational activities, interpersonal relationships, sexual life, and self-esteem. Depression is a frequent consequence of narcolepsy.

The mainstay of treatment is pharmacological although taking naps alleviates excessive sleepiness and refraining from emotion may prevent cataplexy.

Modafinil is the first-line treatment of excessive daytime sleepiness and irresistible episodes of sleep. Methylphenidate is less used today. Based on several large randomized controlled trials showing the activity of sodium oxybate on excessive daytime sleepiness and irresistible episodes of sleep, there is a growing practice in the United States to use it for the later indications. However, adverse effects including nausea, nocturnal enuresis, confusional arousals, malaise, and headache are not exceptional. Sodium oxybate is the only registered treatment for cataplexy. Other treatments are antidepressants, including tricyclics, selective serotonion reuptake inhibitors (SSRIS) and more recently new antidepressants such as venlafaxine or atomoxetine, the later being increasingly used despite few or non-randomized placebo-controlled clinical trials. As for disturbed nocturnal sleep the most widely used treatment is still hypnotics. However, the same randomized controlled trials have shown the activity of sodium oxybate against disturbed nocturnal sleep.

(ii) Narcolepsy without cataplexy

Narcolepsy without cataplexy is described as excessive daytime sleepiness and irresistible episodes of sleep unaccompanied by cataplexy. Automatic behaviour may be present as may hypnagogic hallucinations or sleep paralysis. Nocturnal sleep is usually less disturbed than in narcolepsy with cataplexy. When the subject is young, cataplexy may develop later in the course of the disorder.

Given the risk of overdiagnosing, a positive diagnosis of narcolepsy without cataplexy cannot be done without an all-night polysomnography followed by an MSLT documenting a mean sleep latency of less than 8 min and two or more SOREMPs.

(iii) Narcolepsy due to medical condition

As in the two previous categories, the patient has a complaint of excessive daytime sleepiness occurring almost daily for at least 3 months. However, the distinct feature of this subtype of narcolepsy is the existence of a significant underlying medical or neurological disorder accounting for the excessive daytime sleepiness and/or cataplexy.

According to the *ICSD*-2 the diagnosis of narcolepsy due to medical condition can be made only if one of the following is observed:

- a definite history of cataplexy
- a polysomnographic monitoring followed by an MSLT demonstrating a mean sleep latency of less than 8 min with two or more SOREMPs
- hypocretin-1 levels in the CSF lower than 110 pg/ml, provided the patient is not comatose.

In addition, a consistent chronological link with the presumed underlying disease must be established.

Narcolepsy due to a medical condition is extremely rare.

(b) Idiopathic hypersomnia

In comparison with narcolepsy, which is characterized by well-defined clinical, polysomnographic, and biochemical features, idiopathic hypersomnia is not as well delineated.

Its first description dates back to 1976,[9] almost a century after that of narcolepsy. According to various sleep disorders centres series populations the ratio of idiopathic hypersomnia to narcolepsy with cataplexy would be around 15 per cent.

Idiopathic hypersomnia includes two forms referred to as idiopathic hypersomnia with long sleep time and idiopathic hypersomnia without long sleep time. The first one is remarkable for three symptoms: a complaint of constant or recurrent excessive sleepiness and unwanted naps, usually longer and less irresitible than in narcolepsy, and non-refreshing irrespective of their duration; night sleep is sound, uninterrupted, and prolonged; morning or nap awakening is laborious. Subjects do not awaken to the ringing of a clock, of a telephone, and often rely on their family members who must use vigorous and repeated procedures to wake them up. Even then patients may remain confused, unable to react adequately to external stimuli, a state referred to as "sleep drunkenness" In contrast, idiopathic hypersomnia without long sleep time stands as isolated excessive daytime sleepiness.

The diagnosis of idiopathic hypersomnia is mainly based on clinical features. However, laboratory tests are necessary to rule out other hypersomniac conditions. Polysomnographic monitoring of nocturnal sleep demonstrates normal sleep, except for its prolonged duration in the case of idiopathic hypersomnia with long sleep time. NREM and REM sleep are in normal proportions. There is no SOREMP. Sleep apnoeas and periodic limb movements should theoretically be absent, but may be acceptable in the case of an early onset of idiopathic hypersomnia and of their late occurrence. The MSLT usually demonstrates a mean sleep latency less than 8 min, but is typically longer than in narcolepsy. Fewer than two SOREMPs are present. In cases with normal MSLT findings, several authors have suggested the use of long-term sleep monitoring to document the excessive amount of sleep.

In contrast with narcolepsy, onset of idiopathic hypersomnia is much more progressive. The disorder is generally stable and long lasting. Complications are mostly social and professional.

Treatment of idiopathic hypersomnia relies mainly on modafinil. However, awakening difficulties are hardly improved by this treatment.

(c) Recurrent hypersomnia

The most classic form of recurrent hypersomnia is the Kleine–Levin syndrome.[10] This is an uncommon disorder with roughly 300 cases published in the world literature. Adolescent males are most commonly affected.

The syndrome is characterized by recurrent episodes of hypersomnia associated with behavioural disorders including binge eating (rapid consumption of a large amount of food on a compulsory manner), oversexuality in the form of sexual advances, shamlessly expressed fantasies or masturbation in public, irritability, odd behaviours (like standing on the head, singing loudly, talking in a childish manner), and cognitive disorders, feeling of unreality, confusion, visual or auditory hallucinations. Simultaneous occurrence of all these symptoms is more the exception than the rule however. During the episode the patient may sleep as long as 14 to 18 h per day, waking or getting up only to eat and void.

Body weight gain of a few kilograms can be observed during the episode.

Amnesia of the episode, transient dysphoria, or elation with insomnia for 1 or 2 days, may follow the episode itself. During asymptomatic intervals patients sleep normally and do not experience behavioural or cognitive disorders.

Diagnosis of the Kleine–Levin syndrome is essentially clinical and laboratory tests merely serve to exclude the possibility of rare recurrent hypersomnias of secondary origin, organic, or psychiatric. Of note due to disordered behavioural features, especially hypersexuality, it is not rare that patients are first hospitalized on psychiatric wards and given antipsychotic drugs.

The course is usually benign with episodes lessening in frequency, duration, and severity. Complications are mainly social and occupational.

Treatment of the Kleine–Levin syndrome is not well codified. The effects of stimulant drugs such as modafinil or methylphenidate on the hypersomniac episodes are difficult to assess given that the methods of evaluation are purely subjective and that the episodes vanish spontaneously within a few days. A prophylactic treatment based on mood stabilizers (carbamazepine, lithium carbonate, and more recently valproic acid) deserves to be prescribed in severe cases. Positive results, that is absence of recurrence of the symptoms throughout the period of administration and recurrence with discontinuation, has been reported.

(d) Hypersomnias associated with various medical disorders

(i) Associated with neurological diseases

Hypersomnia may occur in any intracranial pressure syndrome, but it may also result from tumours affecting the diencephalon, specially the ventro-lateral posterior part of the hypothalamus or the peduncular region, with no associated intracranial hypertension. Cases of narcolepsy secondary to brain tumours affecting the hypothalamus or the midbrain region have been reported.

Uni or bilateral paramedian thalamic infarcts and paramedian pedunculo thalamic infarcts are the most typical causes of hypersomnia of vascular origin.

A non-negligible fraction of patients with Parkinson's disease present excessive sleepiness. This is even more the case of patients with multiple system atrophy.

Normal pressure hydrocephalus, Arnold–Chiari malformation, myotonic dystrophy, may also lead to excessive sleepiness.

(ii) Associated with infectious diseases

Intense fatigue and severe excessive sleepiness may develop in the month following Epstein–Barr disease. The same holds true of atypical viral pneumonia, hepatitis B viral infection and the Guillain–Barré syndrome. Hypersomnia tends to go into gradual remission after several months or years.

Disorders of alertness and/or consciousness are found in virtually all patients affected by viral encephalitis.

Human African Trypanosomiasis (sleeping sickness) is a subacute or chronic parasitic disease caused by the inoculation of a protozoan, *Trypanosoma brucei* transmitted by the tsetse fly. It is endemic to certain regions of tropical Africa. The form found in West and Central Africa is due to *Trypanosoma gambiense*. The invasion of the central nervous system is characterized by meningoencephalitis with abnormal sleepiness, headache, trembling, dyskenesia, choreoathetosis, and mood changes. Polysomnography has shown episodes of sleep occurring randomly day and night and SOREMPs.

(iii) Associated with endocrine diseases

Hypothyroidism and acromegaly are the two main sources of hypersomnia, usually due to obstructive sleep apnoea syndrome.

(iv) Post-traumatic hypersomnia

Excessive daytime sleepiness appearing during the year following a head injury may be considered *a priori* as post-traumatic. This typically presents as extended night sleep and episodes of daytime sleep. Brain imaging may reveal lesions affecting the hypothalamic region or brainstem, midbrain or pontine tegmentum, rarely hydrocephalia, or more often the absence of any significant lesions.

The delayed sleep phase syndrome

The major sleep episode is usually delayed 3 to 6 hours relative to conventional sleep-wake times. Affected individuals complain of great difficulty falling asleep at a socially acceptable time, but once sleep ensues, sleep is reported to be normal. Enforced conventional

wake time usually results in morning excessive sleepiness. The disorder has been associated with mental disorders, particularly in adolescents. Schizoid and avoidant personality features are frequently associated with this disorder.

Treatment rests on various approaches including chronotherapy, light exposure, and melatonin therapy.

Conclusion

Hypersomnia deserves to be taken into consideration by the psychiatrist. It may be the consequence of a psychiatric disorder or of its treatment, but it may also be non psychiatric in nature and result in psychiatric symptoms.

Further information

Nofzinger, E.A., Thase, M.E., Reynolds, C.F., *et al.* (1991). Hypersomnia in bipolar depression: a comparison with narcolepsy using the multiple sleep latency test. *The American Journal of Psychiatry*, **148**, 1177–81.

American Academy of Sleep Medicine. (1999). Sleep related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep*, **22**, 667–89.

Dauvilliers, Y., Arnulf, I., and Mignot, E. (2007). Narcolepsy with cataplexy. *Lancet*, **369**, 499–511.

Billiard, M. and Dauvilliers, Y. (2007). Idiopathic hypersomnia. In *Narcolepsy and hypersomnia* (eds. C.L. Bassetti, M. Billiard, and E. Mignot), pp. 77–87. Informa Healthcare, New York.

Arnulf, I., Zeitzer, J.M., File, J., *et al.* (2005). Kleine-Levin syndrome: a systematic review of 186 cases in the literature. *Brain*, **128**, 2763–76.

Regenstein, Q. and Monk, T. (1995). Delayed sleep phase syndrome: a review of its clinical aspects. *The American Journal of Psychiatry*, **152**, 602–8.

References

1. American Academy of Sleep Medicine. (2005). *International classification of sleep disorders: diagnostic and coding manual* (2nd edn). American Academy of Sleep Medicine, Westchester, IL.
2. Ohayon, M.M., Caulet, M., Philip, P., *et al.* (1997). How sleep and mental disorders are related to complaints of daytime sleepiness. *Archives of Internal Medicine*, **157**, 2645–52.
3. Liu, X., Uchiyama, M., Kim, K., *et al.* (2000). Sleep loss and daytime sleepiness in the general adult population of Japan. *Psychiatric Research*, **93**, 1–11.
4. Guilleminault, C., Tilkian, A., and Dement, W.C. (1976). The sleep apnea syndromes. *Annual Review of Medicine*, **27**, 465–84.
5. Young, T., Palta, M., Dempsey, J., *et al.* (1993). The occurrence of sleep disordered breathing among middle-aged adults. *The New England Journal of Medicine*, **328**, 1230–5.
6. Ohayon, M.M. (2003). The effects of breathing-related sleep disorders on mood disturbances in the general population. *Journal of Clinical Psychiatry*, **64**, 1195–2000.
7. Westphal, C. (1877). Eigentümliche mit Einschlafen verbundene Anfälle. *Archives für Psychiatrie und Nervenkrankheiten*, **7**, 631–5.
8. Gelineau, J. (1880). De la narcolepsie. *Gazette des Hôpitaux* (Paris), **55**, 626–8; 635–7.
9. Roth, B. (1976). Narcolepsy and hypersomnia. *Schweizer Archiv für Neurologie und Psychiatrie*, **119**, 31–41.
10. Critchley, M. (1962). Periodic hypersomnia and megaphagia in adolescent males. *Brain*, **85**, 627–56.

4.14.4 Parasomnias

Carlos H. Schenck, and Mark W. Mahowald

In all of us, even in good men, there is a lawless, wild-beast nature which peers out in sleep. –Plato, *The Republic*

Relevance of parasomnias to psychiatrists

Parasomnias are defined as undesirable physical and/or experiential phenomena accompanying sleep that involve skeletal muscle activity (movements, behaviours), autonomic nervous system changes, and/or emotional-perceptual events.[1] Parasomnias can emerge during entry into sleep, within sleep, or during arousals from any stage of sleep; therefore, all of sleep carries a vulnerability for parasomnias.[1] Parasomnias can be objectively diagnosed by means of polysomnography (i.e. the physiologic monitoring of sleep—figures 4.14.4.1, 4.14.4.2), and can be successfully treated in the majority of cases.[2–5] Understanding of the parasomnias, based on polysomnographic documentation, has expanded greatly over the past two decades, as new disorders have been identified, and as known disorders have been recognized to occur more frequently, across a broader age group, and with more serious consequences than previously understood.[1–10] Parasomnias demonstrate how our instinctual behaviours, such as locomotion, feeding, sex, and aggression, can be released during sleep, itself a basic instinct. There are at least eight reasons why parasomnias should be of interest and importance to psychiatrists:

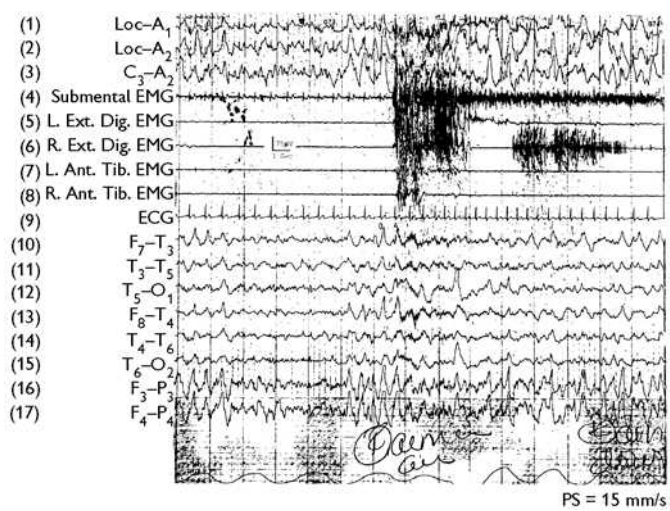


Fig. 4.14.4.1 Polysomnogram of a disordered arousal, with the persistence of sleep, in a 23-year-old man with a history of sleepwalking and sleep terrors. After a behavioural arousal from slow-wave sleep (with arm lifted up and then down), the EEG shows irregular delta and theta activity and superimposed faster frequencies. Immediately preceding the arousal, there is a cluster of three high-amplitude delta waves (channel 3). Electro-occulogram, channels 1. 2; EEG, channels 3, 10–17; EMG, electromyogram. (Reproduced from C.H. Schenck *et al.* Analysis of polysomnographic events surrounding 252 slow-wave sleep arousals in thirty-eight adults with injurious sleepwalking and sleep terrors. *Journal of Clinical Neurophysiology*, **15**, 159–66, copyright 1998, American Clinical Neurophysiology Society)

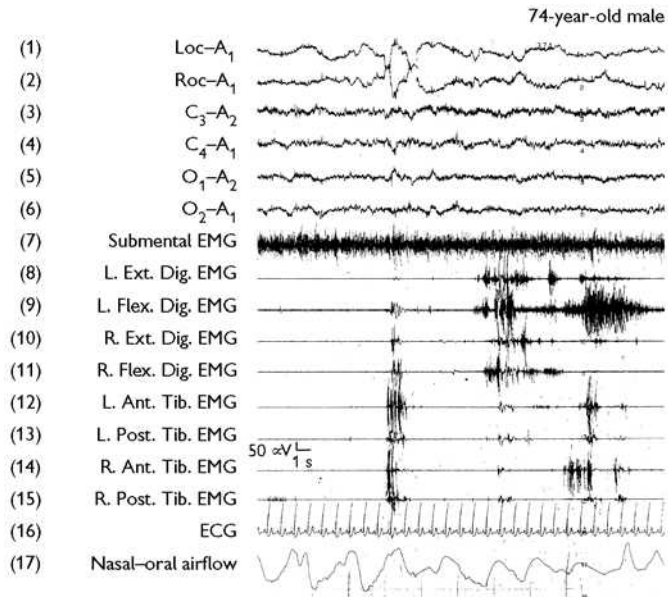


Fig. 4.14.4.2 Polysomnogram of disordered rapid eye movement (REM) sleep in a man with REM sleep behaviour disorder who eventually developed Parkinson's disease. There is complete loss of 'REM-atonia', as the submental electromyogram (EMG) shows continuous muscle tone (channel 7). The appearance of a rapid eye movement (channels 1, 2) signals the onset of excessive muscle twitching in the upper/lower extremity EMGs (channels 8–15). The EEG (channels 3–6) shows the typical low-voltage fast-frequency desynchronized activity of REM sleep. ECG rate (channel 16) remains constant despite generalized muscle twitching, which is a common finding in REM sleep behaviour disorder. Elecro-oculogram: channels 1,2.

1 Parasomnias can be misdiagnosed and inappropriately treated as a psychiatric disorder.

2 Parasomnias can be a direct manifestation of a psychiatric disorder, e.g. dissociative disorder, nocturnal bulimia nervosa.

3 The emergence and/or recurrence of a parasomnia can be triggered by stress.

4 Psychotropic medications can induce the initial emergence of a parasomnia, or aggravate a preexisting parasomnia.

5 Parasomnias can cause psychological distress or can induce or reactivate a psychiatric disorder in the patient or bed partner on account of repeated loss of self-control during sleep and sleep-related injuries.

6 Familiarity with the parasomnias will allow psychiatrists to be more fully aware of the various medical and neurological disorders, and their therapies, that can be associated with disturbed (sleep-related) behaviour and disturbed dreaming.

7 Parasomnias present a special opportunity for interlinking animal basic science research (including parasomnia animal models) with human (sleep) behavioural disorders.

8 Parasomnias carry forensic implications, as exemplified by the newly-recognized entity of 'Parasomnia Pseudo-suicide.' Also, psychiatrists are often asked to render an expert opinion in medicolegal cases pertaining to sleep-related violence.

Classification of parasomnias

Parasomnias can be classified according to whether the signs or symptoms are primary phenomena of sleep itself, or whether they are secondary phenomena derived from various underlying disorders, with sleep facilitating the nocturnal manifestation of these disorders.[6] Table 4.14.4.1 contains such a classification, and provides a context (along with other current sources[1,11]) for the parasomnias to be discussed in this chapter. Parasomnias demonstrate how sleep and wakefulness are not mutually exclusive states. Features of rapid eye movement (REM) sleep, non-REM sleep and wakefulness can occur simultaneously, and with rapid oscillations.[12] Status dissociatus represents the most extreme form of state dissociation.[13]

Clinical evaluation of parasomnias

The evaluation of complex and violent nocturnal behaviours at our centre (Minnesota Regional Sleep Disorders Center) includes the following:[2,14]

1 Clinical sleep-wake interview, with review of medical records, and review of a patient questionnaire that covers sleep-wake, medical, psychiatric and alcohol/chemical use and abuse history, review of systems, family history, and past or current physical, sexual, and emotional abuse.

2 Psychiatric and neurologic interviews and examinations, including psychometric testing.

Table 4.14.4.1 Classification of parasomnias: primary and secondary sleep phenomena

Primary sleep phenomena
Non-REM sleep
Disorders of arousal: sleepwalking/sleep terrors/confusional arousal
REM sleep
REM sleep behaviour disorder (RBD)
Dream anxiety attacks (nightmares)
Miscellaneous (including mixed non-REM/REM sleep)
Parasomnia overlap disorder (sleepwalking/sleep terrors/RBD)
Sleep related eating disorder
Restless legs syndrome/periodic limb movements in sleep
Obstructive sleep apnoea-related parasomnias
Rhythmic movement disorders
Status dissociatus
Bruxism
Secondary sleep phenomena
Central nervous system
Seizures (conventional, unconventional)
Headaches
Psychiatric
Nocturnal dissociative disorders
Nocturnal panic attacks
Nocturnal bulimia nervosa
Post-traumatic stress disorder
Cardiopulmonary (arrhythmias, asthma, etc.)
Gastrointestinal (gastro-oesophageal reflux etc)
Malingering

Modified from Mahowald and Ettinger.[7]

3 Extensive overnight polysomnographic monitoring at a hospital sleep laboratory, with continuous audio-visual recording. Figures 4.14.4.1, 4.14.4.2 depict the polysomnographic montage that includes the electro-oculogram, EEG, chin and four-limb electromyograms, ECG, and nasal-oral airflow (with full respiratory monitoring whenever indicated). Polysomnographic recording speeds of 15 to 30 mm/s are employed in order to detect epileptiform activity. Urine toxicology screening is performed whenever indicated.

4 Daytime multiple sleep latency testing, if there is a complaint or suspicion of excessive daytime sleepiness or fatigue (methods discussed in Narcolepsy chapter).

Causes of sleep related injury

A report on a series of 100 consecutive adults presenting to a multidisciplinary sleep disorders centre on account of sleep-related injury identified five causes:[2]

1 Sleepwalking/sleep terrors (M:F 3:2; mean age of onset, 5 years);

2 REM sleep behaviour disorder (predominantly male; mean age of onset, 57 years)

3 Dissociative disorders (Mostly female mean age of onset, 21 years)

4 Nocturnal seizures (uncommon)

5 Obstructive sleep apnoea/periodic limb movements (uncommon).

The sleep-related injuries included ecchymoses, lacerations, and fractures in 95 per cent, 30 per cent, and 9 per cent of patients respectively.

Non-REM sleep parasomnias: sleepwalking and sleep terrors

The polysomnographic correlates of sleepwalking and sleep terrors were first identified in the 1960s and 1970s by Gastaut and Broughton,[15, 16] Kales *et al.*,[17] and Fisher *et al.*[18] from France, Canada, and the United States. Sleepwalking and sleep terrors are classified as 'disorders of arousal,' and typically arise from delta non-REM (slow wave) sleep and usually affect children, but adults can also be afflicted, and suffer from sleep-related injuries and adverse social consequences.[2,9,10,19–24]

Clinical findings

Sleepwalking (SW) is characterized by complex, automatic behaviours, such as aimlessly wandering about, nonsensically carrying objects from one place to another, rearranging furniture, eating inappropriately, urinating in cupboards, going outdoors, and on rare occasion, driving a car.[1,25,26] The eyes are usually wide open and have a glassy stare, and there may be some mumbling. However, communication with a sleepwalker is usually poor or impossible. Frenzied or aggressive behaviour, the wielding of weapons (knives, guns), or the calm suspension of judgement (e.g. leaving via a bedroom window, wandering far outdoors) can result in inadvertent injury or death to self or others. Homicidal sleepwalking can occur.[25]

Sleepwalking episodes usually emerge 15–120 minutes after sleep onset, but can occur throughout the entire sleep period in adults. The duration of each episode can vary widely. The following is a wife's description.[2]

'He seems to have the strength of 10 men and shoots straight up from bed onto his feet in one motion. He's landed clear across the room on many occasions and has pulled down curtains (bending the rods), upset lamps, and so forth. He's grabbed me and pulled on me, hurting my arms, because he's usually dreaming that he's getting me out of danger . . . He's landed on the floor so hard that he's injured his own body . . . There are low windows right beside our bed and I'm afraid he'll go through them some night.'

Sleep terrors (ST) are characterized by sudden, loud, terrified screaming and prominent autonomic nervous system activation (tachycardia, tachypnea, diaphoresis, mydriasis) that usually appears early in the sleep period, although episodes in adults can occur at any time of the night. The individual may sit up rapidly while screaming, and engage in frenzied activity, such as bolting out of bed, and becoming injured.

Childhood sleepwalking and sleep terrors are characterized by complete amnesia for the events. In adult SW and ST, there can be subsequent recall of the behavioural episode, and also recall of dream-like mentation that usually involves being threatened by imminent danger.[2,27] The distinction between ST and agitated SW in adults is often blurred, with both states being admixed in response to a perceived threat.[24]

The prevalence of SW has been estimated to be as high as 17 per cent in childhood (peaking at age 4–8), and recent data indicate a higher prevalence in adults (4 per cent) than previously recognized.[28–30] The prevalence of ST in children can be greater than 6 per cent and greater than 2 per cent in adults.[1] A familial-genetic basis for SW and ST has been well-established.[30,31] Non-injurious SW does not have a gender preference,[1] although injurious SW appears to be more male-predominant.[2] Sleep terrors do not have a gender preference.[1] 'Confusional arousals' comprise another category of 'disorder of arousal,' and represent partial manifestations of sleepwalking and sleep terrors in which aggression and sexual impulses can be released.[1,10]

Polysomnographic findings

Sleepwalking/sleep terrors episodes arise abruptly during arousals from delta non-REM sleep.[1,15–18] In a systematic study of 38 adults with injurious SW/ST,[24] three postarousal EEG patterns were detected: diffuse, rhythmic delta activity; diffuse delta and theta activity intermixed with alpha and beta activity; and prominent alpha and beta activity. Thus, the postarousal EEG can show the complete persistence of sleep, the admixture of sleep and wakefulness, or complete wakefulness. Figure 4.14.4.1 shows the polysomnogram of a disordered arousal from slow-wave sleep.

Although the sleep architecture (i.e. distribution of sleep stages) is usually normal in SW and ST, the 'micro-structure' of non-REM sleep in adult SW/ST can be perturbed, with increased micro-arousals and increased rate of the 'cyclic alternating EEG pattern'.[1,2,21,23]

Association with medical and psychiatric risk factors

Sleepwalking and sleep terrors may be triggered by sleep deprivation, febrile illness, alcohol use or abuse, pregnancy, menstruation, obstructive sleep apnea, periodic limb movements, nocturnal

seizures, medical and neurological disorders, and psychotropic medications—especially zolpidem, lithium carbonate and anticholinergic agents.[1]

A strong association between sleepwalking/sleep terrors and psychopathology in adults was suggested by an early literature, but polysomnograpic monitoring was not conducted, and there were considerable methodological problems. The recent literature reporting PSG-confirmed cases has indicated that most adult cases are not closely associated with a psychiatric disorder, although stress can play a promoting role[2,19,29,22,32]. Genetic-constitutional factors appear to be predominant in adult and childhood sleepwalking/sleep terrors.[1]

Treatment

Treatment (especially in childhood SW/ST) is usually not necessary, other than identifying and minimizing any identified risk factors, and safety measure to avoid accidental injury. In cases involving sleep-related injury, pharmacological treatment is necessary and can be life-saving. A benzodiazepine, such as clonazepam (0.25–1.5 mg) taken 1 hour before sleep onset is usually effective. Alprazolam, diazepam, imipramine, and paroxetine can also be used. Teaching a patient self-hypnosis can be effective in milder cases of either adult or childhood SW/ST.[33] Treatment of any concurrent psychiatric disorder does not usually control sleepwalking/sleep terrors.[2,22] Attempts to waken the patient may cause confusion and distress.

REM sleep behaviour disorder (RBD)

Although various aspects of RBD have been identified by European, Japanese and American investigators since 1966,[34] RBD was not formally recognized and named until 1986–1987,[35,36] and it is incorporated within the international classification of sleep disorders.[1] A typical clinical presentation of RBD is contained in the description of the index case:[35]

> 'A 67-year-old dextral man was referred because of violent behavior during sleep . . . 4 years before referral . . . he experienced the first 'physically moving dream' several hours after sleep onset; he found himself out of bed attempting to carry out a dream. This episode signaled the onset of an increasingly frequent and progressively severe sleep disorder; he would punch and kick his wife, fall out of bed, stagger about the room, crash into objects, and injure himself . . . his wife began to sleep in another room 2 years before referral. They remain happily married, believing that these nocturnal behaviors are out of his control and discordant with his waking personality.'

Mammalian REM sleep, REM atonia, and paradox lost

REM sleep in mammals involves a highly energized state of brain activity, with both tonic (i.e. continuous) and phasic (i.e. intermittent) activations occurring across a spectrum of physiologic parameters.[36] REM sleep has two major synonyms: 'active sleep,' because of the high level of brain activity during REM sleep, and 'paradoxical sleep,' because of the nearly complete suppression of muscle tone despite the high level of brain activity. This generalized skeletal muscle atonia ('REM-atonia') is one of the three defining features of mammalian REM sleep, besides rapid eye movements and a desynchronized EEG. The loss of the customary paradox of REM sleep in RBD bears serious clinical consequences: 'paradox lost' means loss of safe sleep.[9]

Animal model of RBD

In 1965, Jouvet and Delorme reported from France that experimental pontine lesions in cats caused permanent loss of REM-atonia, and the cats displayed attack and exploratory behaviours during REM sleep that resembled dream-enactment (oneirism). This experiment established the first animal model of RBD.[34]

Clinical and polysomnographic findings

Between 1982 and 1991, 96 patients at our centre were diagnosed with chronic RBD. (There is an acute form of RBD that can emerge during withdrawal from ethanol or sedative-hypnotic abuse and with anti-cholinergic and other drug intoxication states.[36] Data on this series[37] are contained in Table 2, and are concordant with the published world literature.[34,36] The older male predominance in RBD is striking, although females and virtually all age groups are represented. Approximately half of RBD cases are closely associated with neurological disorders, predominantly neurodegenerative disorders (especially parkinsonism), narcolepsy and stroke. In fact, RBD may be the first sign of a parkinsonian disorder whose other (classic) manifestations may not emerge until several years or even decades after the onset of RBD.[36,38,39] Thus, routine neurological evaluations are indicated in the long-term management of RBD. The prevalence of RBD is estimated to be from 0.38 per cent to 0.5 per cent.[1] The course is usually progressive; spontaneous

Table 4.14.4.2 Findings in 96 patients with chronic REM sleep behaviour disorder (RBD)

Categories	Percentage (*N*)	Comments
Gender		
Male	87.5 (84)	Mean age of RBD onset (*N* = 90): 52 years (range 9–81)
Female	12.5 (12)	Mean age at polysomnography: 58 years (range 10–83)
Chief complaint		
Sleep injury	79.2 (76)	Ecchymoses (76); lacerations (32); fractures (7)
Sleep disruption	20.8 (20)	
Altered dream process and content	87.5 (84)	More vivid, intense, action filled, violent (reported as severe nightmares)
Dream-enacting behaviours	87.5 (84)	Talking, laughing, yelling, swearing, gesturing, reaching, grabbing, arm flailing, punching, kicking, sitting, jumping out of bed, crawling, running
Clonazepam treatment		
Efficacy (*N* = 67)		
Complete	79.1 (44/67)	Rapid control of problematic sleep behaviours *and* altered dreams, sustained for up to 9 years
Partial	11.9 (8/67)	
Total	91.0 (61/67)	

Modified from Schenck *et al.*[36]

remissions are very rare. Patients with RBD usually have calm and pleasant personalities, and do not display irritability or anger while awake, even though their dreams are highly aggressive.[40] Figure 4.14.4.2 depicts a typical polysomnogram of RBD with attempted dream-enactment.

Association of RBD with psychiatric disorders and Stress

Psychiatric disorders are rarely associated with RBD.[34,41] Fluoxetine treatment of obsessive compulsive disorder,[42] or cessation of use or abuse of REM-suppressing agents (viz. ethanol, amphetamine, cocaine, imipramine) can trigger RBD.[36,43] Tricyclic antidepressants, selective serotonin reuptake inhibitors, venlafaxine, mirtazapine, and monoamine oxidase inhibitors can induce or aggravate RBD. In four cases, major stress involving divorce, automobile accident, sea disaster, or public humiliation triggered RBD.[2,34,43]

Diagnosis of RBD

The diagnostic criteria of RBD are as follows:[1,36]

1 Polysomnographic abnormality during REM sleep: elevated submental electromyographic tone and/or excessive phasic submental and/or limb electromyographic twitching.

2 Documentation of abnormal REM sleep behaviours during polysomnographic studies, or a history of injurious or disruptive sleep behaviours.

3 Absence of EEG epileptiform activity during REM sleep.

Treatment of RBD

Clonazepam is remarkably effective in controlling both the behavioural and the dream-disordered components of RBD, at a usual bedtime dose of 0.5 to 1.0 mg. The long-term efficacy and safety of chronic, nightly clonazepam treatment of RBD has been established.[3] Other treatments include melatonin, pramipexole, etc.[36] Maximizing the safety of the sleeping environment should always be encouraged.

Parasomnia overlap disorder: sleepwalking/ sleep terrors/RBD

A group of 33 patients has been reported with polysomnogram-documented sleepwalking, sleep terrors, and RBD.[8] The mean age was 34 years, the mean age of parasomnia onset was 15 years (range: 1–66), and 70 per cent were male. An idiopathic subgroup (N=22) had a significantly earlier mean age of parasomnia onset (9 years) than a symptomatic subgroup (27) whose parasomnia began with either a neurologic disorder (N=6), nocturnal paroxysmal atrial fibrillation (N=1), post-traumatic stress disorder/ major depression (N=1), chronic ethanol/amphetamine abuse and withdrawal (N=1), or mixed disorders (schizophrenia; brain trauma; substance abuse [N=2]). The rate of psychiatric disorders was not elevated; group scores on various psychometric tests were not elevated. Forty-five percent (N=15) had previously received psychologic or psychiatric therapy for their parasomnia, without benefit. Treatment, usually with clonazepam, was effective for most patients. The natural history of this disorder is not yet known. Other cases of Parasomnia Overlap Disorder have been reported.[34,36]

Sleep related eating disorder

The 'night-eating syndrome' was first reported in 1955, and featured abnormal eating during full wakefulness in insomniac patients. Over the next 35 years, abnormal nocturnal eating received scant attention in the literature, until a series of 19 cases (expanded to 38 cases) with polysomnographic data was published on the new entity of sleep related eating disorder,[1,4,44] which featured abnormal eating during partial arousals from sleep in patients with other parasomnias or those who were idiopathic. Reports from various countries have now been published.[45–49]

Clinical findings

Pathological sleep-related eating has characteristic features, as identified in two separate series of adult cases.[4,44,48] Neither daytime binge-eating nor obsessive-compulsive disorder was diagnosed in any patient. Sleep-related eating was not associated with either the onset or the course of a psychiatric disorder. Patients do not usually experience hunger or thirst during their 'driven' nocturnal eating and drinking. A typical behavioural sequence consists of 'automatic' arising from bed, and going straight to the kitchen with a compulsive 'out of control' urge to eat. Alcohol is almost never consumed, even in former alcohol abusers. Thick substances, such as milkshakes, peanut butter, and brownies, are preferentially consumed at night. Purging does not take place, either at night or in the morning. Sleep-related eating is usually invariant, and is not influenced by day of the week or sleeping away from home. More than 40 per cent of patients in one series were overweight from the nocturnal eating, according to standardized body mass index criteria.[4]

Sleep related eating is most commonly associated with sleepwalking, but also with restless legs syndrome, obstructive sleep apnoea, narcolepsy, zolpidem/triazolam/midazolam/other psychotropic medication use or abuse, cessation of cigarette smoking, cessation of alcohol, opiate and cocaine abuse; cessation of cigarette smoking; stress (particularly involving separation anxiety), and various medical/neurological disorders (e.g. autoimmune hepatitis, encephalitis, migraines). Rarely, anorexia or bulimia nervosa or a dissociative disorder can be associated with sleep-related eating.[45]

Prescribing a monoamine oxidase inhibitor to a patient with sleep related eating disorder can be hazardous, since indiscriminate food consumption at night could jeopardize the mandatory dietary restrictions of MAOI treatment.

Treatment

Treatment, with one notable exception, is primarily directed at controlling the underlying sleep disorder. For cases associated with restless legs syndrome, therapy with dopaminergics (at times supplemented with an opiate such as codeine) is usually effective; likewise, therapy of obstructive sleep apnoea with nasal continuous positive airway pressure is often effective in controlling the sleep related eating. In contrast, in cases that are idiopathic and in those cases associated with sleepwalking, therapy with the anticonvulsant topiramate (that suppresses overeating urges) was effective in two-thirds of cases in two reported series.[5,50]

Sleep related eating disorder shares many features in common with the 'night eating syndrome' in adults,[49,51] although there are usually major differences in regards to level of consciousness during nocturnal eating, association with other sleep disorders,

and response to therapy. It is likely that these two conditions sit at opposite poles of a spectrum of abnormal nocturnal eating.

Sleep related dissociative disorders

A nocturnal dissociative disorder with polysomnographic monitoring was first reported in 1976 by Rice and Fisher in a man with daytime and night-time fugues that began after his father's death.[52] Another polysomnogram-documented case was described in a woman with a history of being physically and sexually assaulted.[1] A series of eight cases was reported from our centre in 1989.[53]

Clinical manifestations

The onset may be sudden or gradual, and the course is chronic. There usually is a history of repeated physical and sexual abuse in childhood and/or adulthood. In a series of eight cases, seven were female who also had daytime states of dissociation with self-mutilating behaviours, such as genital cutting, self-burning, and punching through windows.[53] One male patient had exclusively nocturnal episodes in which he acted like a jungle cat. Patients may report sleep phobia as a consequence of bed-related and sleep-related sexual and physical abuse.

A typical spell during polysomnographic monitoring involves complex and lengthy behaviours that emerge during well-established EEG wakefulness, after a prior episode of sleep.[53] The nocturnal episodes appear to be re-enactments of previous assaults.

Treatment

Treatment involves long-term therapy of the dissociative disorders and of associated psychiatric disorders, usually initiated in a specialized in-patient setting. Bedtime administration of benzodiazepines may aggravate a nocturnal dissociative disorder.

Restless legs syndrome (RLS)

RLS, first described in 1945 by Ekbom from Sweden, is a chronic disorder that often results in severe insomnia, and can be incapacitating.[1,54] It is characterized by unpleasant and at times painful sensations in the lower extremities that emerge during periods of inactivity, particularly during the transition from wakefulness to sleep. These abnormal sensations are relieved by movement or stimulation of the legs, such as walking, stomping the feet, rubbing or squeezing the legs, taking hot showers, or applying hot packs or ointments to the legs.

Both the RLS and the therapeutic interventions just described are incompatible with successful entry into sleep. The more severely affected individuals cannot easily sit still while watching television or a film, or during protracted plane or train journeys. RLS is quite common, affecting up to 10 per cent of the general population, and tends to become more prevalent with increasing age. It affects females more than males. Childhood cases, though rare, at times may masquerade as 'attention deficit disorder with hyperactivity,' or as 'growing pains.'

Most RLS cases are idiopathic with a strong familial basis. Caffeine, fatigue, or stress may worsen the symptoms. There is no evidence that RLS is related to psychiatric disorders, although the symptoms of RLS may suggest a primary anxiety disorder and through impairment of the duration and quality of sleep it can affect daytime mood and behaviour. Also, neuroleptic-induced akathisia can mimic RLS. Secondary RLS can be associated with peripheral neuropathies, renal disease, and psychotropic medications (especially the SSRIs, venlafaxine, and tricyclic antidepressants).

Nearly all patients with RLS have 'periodic limb movements' (PLMs) of non-REM sleep (formerly called 'nocturnal myoclonus'). PLMs are characterized by periodic (every 15–40 sec) movements of the legs, viz. slow dorsi-flexion, which may or may not be associated with arousals. Excessive arousals during sleep from any cause, including PLMs, may result in daytime fatigue or sleepiness.

RLS can usually be diagnosed by history alone, and polysomnographic monitoring is usually not indicated. The syndrome is one of the major organic causes of insomnia, and commonly responds to treatment that includes use of dopaminergics (e.g. pramipexole, ropinerole, levodopa), opiates, and benzodiazepines (e.g. clonazepam). Combinations of these drugs are often necessary. Since RLS is a chronic disorder that often worsens with advancing age, chronic long-term treatment is usually warranted.

Differential diagnosis of dream-enacting behaviours and other parasomnias

A history of dream-enacting behaviours does not automatically implicate RBD. Other diagnoses include sleepwalking/sleep terrors;[2,9] obstructive sleep apnoea, with apnoea-induced arousals from REM sleep being associated with violent behaviours and vivid REM-related dreams;[2,55] nocturnal complex seizures, with the 'dreams' being the seizure equivalents;[34] sleep related dissociative disorders, with the 'dreams' being wakeful memories of past abuse;[53] intoxication states; and malingering.

Other parasomnias of interest to psychiatrists include sexual parasomnias ('sleepsex', 'sexsomnia'),[10] nocturnal panic attacks, which arise from Stages II and III non-REM sleep,[56] sleep related trichotillomania, nocturnal frontal lobe epilepsy,[11] and rhythmic movement disorders of non-REM and REM sleep, including head-banging and body rocking. Finally, expanded knowledge on the parasomnias has allowed for the recognition of a new medical-legal entity called 'parasomnia pseudo-suicide'.[58] This term refers to the unfortunate, but unintentional, fatal consequence of complex, sleep-related behaviours that may be erroneously attributed to suicide.[1]

Forensic guidelines for psychiatrists

The legal implications of automatic behaviour have long been discussed and debated. With regard to sleep-related automatic behaviours, the objective identification of a sleep disorder does not establish causality for any given deed. Guidelines have been developed to assist in determining the role of a sleep disorder in an act of violence.[57]

- There should be reason (by history or by formal sleep laboratory evaluation) to suspect a bona fide sleep disorder. Similar episodes, with benign or morbid outcome, should have occurred previously. (Note that disorders of arousal may begin in adulthood).
- The duration of the action is usually brief (minutes).
- The behaviour is usually abrupt, immediate, impulsive, and senseless—without apparent motivation. Although ostensibly

purposeful, it is completely inappropriate to the total situation, out of (waking) character for the individual, and without evidence of premeditation.

- The victim is someone who merely happened to be present and who may have been the stimulus for the arousal.
- Immediately after return to consciousness, there is perplexity or horror, without an attempt to escape, conceal, or cover up the action. There is evidence of lack of awareness on the part of the individual during the event.
- There is usually some degree of amnesia for the event, however, this amnesia may not be complete.
- In the case of ST or SW or sleep drunkenness, the act may (a) occur on awakening, (rarely immediately upon falling asleep) and usually at least 1 hour after sleep onset; (b) occur on attempts to awaken the subject; and (c) have been potentiated by alcohol ingestion, sedative or hypnotic administration or prior sleep deprivation.

The American Academy of Neurology and other professional organizations have developed guidelines for expert witnesses, with the most stringent being as follows.

- The expert should be willing to submit his or her testimony for peer review.
- The expert must be impartial and be willing to prepare his or her testimony for use by either or both the plaintiff or defendant. Ideally, the expert should assume the role of 'friend of the court.'

Further information

Schenck, C.H., Mahowald, M.W. Parasommias (2008). Associated with sleep-disordered breathing and its therphy, including sexsomnia as a recently recognized parasomnia. *Somnology*, **12**, 38–49.

Howell, M.J., Schenck, C.H., Crow, S.J. (2008). A review of night-time eating disorders. Sleep Medicine Reviews; doi: 10.1016/j.smrv.2008.07.005.

Schenck, C.H. (2005). *Paradox lost: midnight in the battleground of sleep and dreams*. Extreme-Nights, LLC, Minneapolis, Minnesota www.parasomnias-rbd.com

"Sleep Runners: The Stories Behind Everyday Paramnias—Deluxe Academic Edition" (3 DVDs, including 50 minutes of indexed sleep lab parasomnia footage; and CD-ROM), Slow-Wave Films, LLC, St. Paul, Minnesota, 2007. www.sleeprunners.com

www.parasomnias-rbd.com

References

1 American Academy of Sleep Medicine (2005). *International classifi cation of sleep disorders:Diagnostic and coding manual* (2nd edn.). Westchester, Ill:*American Academy of Sleep Medicine*.

2. Schenck, C.H., Milner, D.M., Hurwitz, T.D., *et al.* (1989). A polysomnographic and clinical report on sleep-related injury in 100 adult patients. *American Journal of Psychiatry*, **146**, 1166–73.
3. Schenck, C.H., and Mahowald, M.W. (1996b). Long-term, nightly benzodiazepine treatment of injurious parasomnias and other disorders of disrupted nocturnal sleep in 170 adults. *American Journal of Medicine*, **100**, 548–54.
4. Schenck, C.H., Hurwitz, T.D., O'Connor, K.A., *et al.* (1993). Additional categories of sleep-related eating disorders and the current status of treatment. *Sleep*, **16**, 457–66.
5. Winkelman J.W. (2006). Efficacy and tolerability of open-label topiramate in the treatment of sleep-related eating disorders: an open-label, retrospective case series. *Journal of Clinical Psychiatry*, **67**(11), 1729–34.
6. Mahowald, M.W., and Ettinger, M.G. (1990). Things that go bump in the night: the parasomnias revisited. *Journal of Clinical Neurophysiology*, **7**, 119–43.
7. Ohayon M.M., *et al.* (1997) Violent behavior during sleep. *Journal of Clinical Psychiatry*, **58,** 369–76.
8. Schenck, C.H., Boyd, J.L., and Mahowald, M.W. (1997). A parasomnia overlap disorder involving sleepwalking, sleep terrors, and REM sleep behavior disorder in 33 polysomnographically confirmed cases. *Sleep,* **20**, 972–81.
9. Schenck, C.H. (2005). *Paradox Lost: Midnight In The Battleground Of Sleep And Dreams*. Minneapolis, Minnesota, Extreme-Nights, LLC.
10. Schenck, C.H., Arnulf, I., Mahowald, M.W. (2007). Sleep and sex: what can go wrong? A review of the literature on sleep disorders and abnormal sexual behaviors and experiences. *Sleep*, **30**, 683–702.
11. Kryger, M.H., Roth, T., and Dement, W.C. (2005). *Principles and Practice of Sleep Medicine* (4th edn.). Elsevier: Philadelphia, PA.
12. Mahowald, M.W. and Schenck, C.H. (1992). Dissociated states of wakefulness and sleep. *Neurology*, **42**(Suppl. 6), 44.
13. Mahowald, M.W. and Schenck, C.H. (1992). Status dissociatus—a perspective on states of being. *Sleep*, **14,** 69.
14. Schenck, C.H., Mahowald, M.W. (2005). Rapid Eye Movement Sleep Parasomnias. *Neurologic Clinics*, **23**: 1107–26.
15. Gastaut, H. and Broughton, R. (1965). A clinical and polygraphic study of episodic phenomena during sleep. *Recent Advances in Biological Psychiatry*, **7**, 197–222.
16. Broughton, R. (1968). Disorders of sleep: disorders of arousal? Science, **59**, 1070–8.
17. Kales, A., Jacobson, A., Paulson, J., *et al.* (1966). Somnambulism: psychophysiological correlates. I. All-night EEG studies. *Archives of General Psychiatry*, **14**, 586–94.
18. Fisher, C., Kahn, E., Edwards, A., *et al.* (1973). A psychophysiological study of nightmares and night terrors. I. Physiological aspects of the stage 4 night terror. *Journal of Nervous and Mental Disease*, **157**, 75–98.
19. Kavey, N.B., Whyte, J., Resor, S.R., *et al.* (1990). Somnambulism in adults. *Neurology*, **49**, 749–52.
20. Crisp, A.H., Matthews, B.M., Oakey, M., *et al.* (1990). Sleepwalking, night terrors, and consciousness. *British Medical Journal*, **300**, 360–2.
21. Blatt, I., Peled, R., Gadoth, N., *et al.* (1991). The value of sleep recording in evaluating somnambulism in young adults. *Electroencephalography and Clinical Neurophysiology*, **78**, 407–12.
22. Llorente, M.D., Currier, M.B., Norman, S.E., *et al.* (1992). Night terrors in adults: phenomenology and relationship to psychopathology. *Journal of Clinical Psychiatry*, **53**, 392–4.
23. Zucconi M, *et al.* (1995) Arousal fl uctuations in non-rapid eye movement parasomnias: the role of cyclic alternating pattern as a measure of sleep instability. *Journal of Clinical Neurophysiology*, **12**:147–54.
24. Schenck, C.H., Pareja, J.A., Patterson, A.L., *et al.* (1998). Analysis of polysomnographic events surrounding 252 slow-wave sleep arousals in thirty-eight adults with injurious sleepwalking and sleep terrors. *Journal of Clinical Neurophysiology*, **15**, 159–66.
25. Broughton, R., Billings, R., Cartwright, R., *et al.* (1994). Homicidal somnambulism: a case report. *Sleep*, **17**, 253–64.
26. Schenck, C.H., and Mahowald, M.W. (1995). A polysomnographically documented case of adult somnambulism with long-distance automobile driving and frequent nocturnal violence: parasomnia with continuing danger as a non-insane automatism? *Sleep*, **18**, 765–72.
27. Kavey, N.B. and Whyte, J. (1993). Somnambulism associated with hallucinations.

28. Bixler, E.O., Kales, A., Soldatos, C.R., *et al.* (1979). Prevalence of sleep disorders in the Los Angeles metropolitan area. *American Journal of Psychiatry*, **136**, 1257–62.
29. Klackenberg, G. (1982). Somnambulism in childhood—prevalence, course and behavior correlates. A prospective longitudinal study (6–16 years). *Acta Paediatrica*, **71**, 495–9.
30. Hublin, C., Kaprio, J., Partinen, M., *et al.* (1997). Prevalence and genetics of sleepwalking: a population-based twin study. *Neurology*, **48**, 177–81.
31. Kales, A., Soldatos, C.R., Bixler, E.O., *et al.* (1980). Hereditary factors in sleepwalking and night terrors. *British Journal of Psychiatry*, **137**, 111–18.
32. Crisp. A.H. (1996). The sleepwalking/night terrors syndrome in adults. *Postgraduate Medical Journal*, **72**, 599–604.
33. Hurwitz, T.D., Mahowald, M.W., Schenck, C.H., *et al.* (1991). A retrospective outcome study and review of hypnosis as treatment of adults with sleepwalking and sleep terror. *Journal of Nervous and Mental Disease*, **179**, 228–33.
34. Schenck, C.H., Mahowald, M.W. (2002). REM Sleep Behaviour Disorder: Clinical, Developmental, and Neuroscience Perspectives 16 Years After Its Formal Identification in Sleep, *Sleep,* **25**,120–38.
35. Schenck, C.H., Bundlie, S.R., Ettinger, M.G., *et al.* (1986). Chronic behavioral disorders of human REM sleep: a new category of parasomnia. *Sleep*, **9**, 293–308.
36. Mahowald, M.W., Schenck, C.H. (2005). REM Sleep Parasomnias. In *Principles and Practice of Sleep Medicine* (4th edn.) (eds. M.H. Kryger, T. Roth, W.C. Dement,), pp. 897–916, Elsevier Saunders, Philadelphia, Pennsylvania.
37. Schenck, C.H., Hurwitz, T.D., and Mahowald, M.W. (1993). REM sleep behaviour disorder: an update on a series of 96 patients and a review of the world literature. *Journal of Sleep Research*, **2**, 224.
38. Gagnon J-F, Postuma R.B., Mazza, S., *et al.* (2006). Rapid-eye-movement sleep behavior disorder and neurodegenerative diseases. *Lancet Neurology*, **5**, 424–32.
39. Iranzo, A., Molinuevo, J.L., Santamaria, J., *et al.* (2006) Rapid-eye-movement sleep behavior disorder as an early marker for a neurodegenerative disorder: a descriptive study. *Lancet Neurology*, DOI: 10.1016/S1474-4422(06)70476–8.
40. Fantini, M.L., Corona, A., Clerici, S., *et al.* (2005) Aggressive dream content without daytime aggressiveness in REM sleep behavior disorder. *Neurology*, **65**, 1010-15.
41. Schenck, C.H., and Mahowald, M.W. (1990). A polysomnographic, neurologic, psychiatric and clinical outcome report on 70 consecutive cases with REM sleep behavior disorder (RBD): sustained clonazepam efficacy in 89.5 per cent of 57 treated patients. *Cleveland Clinical Journal of Medical*, **57**(Suppl), 10–24.
42. Schenck, C.H., Mahowald, M.W., Kim, S.W., *et al.* (1992). Prominent eye movements during NREM sleep and REM sleep behavior disorder associated with fl uoxetine treatment of depression and obsessive-compulsive disorder. *Sleep*, **15**:226–35.
43. Schenck, C.H., Hurwitz, T.D., Mahowald M.W. (1988). REM sleep behaviour disorder. *American Journal of Psychiatry*, **145,** 652.
44. Schenck, C.H., Hurwitz, T.D., Bundlie S.R., *et al.* (1991). Sleep-related eating disorders: polysomnographic correlates of a heterogeneous syndrome distinct from daytime eating disorders. *Sleep,* **14**, 419–31.
45. Schenck, C.H., and Mahowald, M.W. (1994). Review of nocturnal sleep-related eating disorders. *International Journal of Eating Disorders*, **15**, 343–56.
46. Spaggiari, M.C., Granella, F., Parrino, L., (1994). Nocturnal eating syndrome in adults. *Sleep*, **17**, 339–44.
47. Manni, R., Ratti, M.T., and Tartara, A. (1997). Nocturnal eating: prevalence and features in 120 insomniac referrals. *Sleep*, **20**, 734–8.
48. Winkelman, J.W. (1998). Clinical and polysomnographic features of sleep-related eating disorder. *Journal of Clinical Psychiatry*, **59**, 14–19.
49. Vetrugno, R., Manconi, M., Ferini-Strambi, L., *et al.* (2006). Nocturnal eating: sleep-related eating disorder or night eating syndrome? A videopolysomnographic study. *Sleep*, **29**: 949–54.
50. Schenck, C.H., and Mahowald, M.W. (2006). Topiramate therapy of sleep related eating disorder. *Sleep*, **29**:A268.
51. Schenck, C.H. (2006). Journal Search And Commentary: A Study of Circadian Eating and Sleeping Patterns In Night Eating Syndrome (NES) Points The Way To Future Studies On NES And Sleep-Related Eating Disorder. *Sleep Medicine*, **7**, 653–6.
52. Rice, E. and Fisher, C. (1976). Fugue states in sleep and wakefulness: a psychophysiological study. *Journal of Nervous and Mental Disease*, **163**, 79–87.
53. Schenck, C.H., Milner, D.M., Hurwitz, T.D., *et al.* (1989b). Dissociative disorders presenting as somnambulism: polysomnographic, video and clinical documentation (8 cases). *Dissociation*, **2**, 194–204.
54. Montplaisir, J., Allen, R.P., Walters, A.S., *et al.* (2005). Restless legs syndrome and periodic limb movements during sleep. In *Principles and Practice of Sleep Medicine* (4th edn.), (eds. M.H. Kryger, T. Roth, W.C. Dement), pp. 839–849. Elsevier Saunders, Philadelphia, Pennsylvania.
55. Iranzo, A. and Santamaria, J. (2005). Severe obstructive sleep apnea/hypopnea mimicking REM sleep behavior disorder. *Sleep*, **28**: 203–06.
56. Shapiro, C.M. and Sloan, E.P. (1998). Nocturnal panic—an underrecognized entity. *Journal of Psychosomatic Research*, **44**, 21–3.
57. Mahowald, M.W. and Schenck, C.H. (1999). Sleep-related violence and forensic medicine issues. In Sleep disorders medicine: basic science, technical considerations, and clinical aspects (2nd edn) (ed. S. Chokroverty), pp. 729–39. Butterworth Heinemann, Boston, MA.
58. Mahowald, M.W., Schenck, C.H., Goldner, M., *et al.* (2003). Parasomnia Pseudo- Suicide. *Journal of Forensic Sciences*, **48**: 1158–62.

4.15

Suicide

Contents

4.15.1 Epidemiology and causes of suicide

Jouko K. Lonnqvist

Definition of suicide and the reliability of suicide statistics

Suicidal behaviour or suicidality can be conceptualized as a continuum ranging from suicidal ideation and communications to suicide attempts and completed suicide. A developmental process which leads to suicidal ideation, suicidal communication, self-destructive behaviour, in some cases even to suicide, and its consequences to the survivors is often referred to as a suicidal process. There is no single unanimously accepted definition of suicide, although in most proposed definitions it is considered as a fatal act of self-injury (self-harm) undertaken with more or less conscious self-destructive intent, however vague and ambiguous. Since the deceased cannot testify as to his or her intent, the conclusions about this must be drawn by inference. The evidence required for this inference depends on many factors, for example the mode of death, the use of autopsy, age, gender, social and occupational status, and the social stigma of suicide in the person's culture. The assessment of suicide intent is always based on a balance of probabilities.

Besides the conceptual problems, there are differences in operational definitions of suicidal behaviour which may lead to lack of uniformity of case definition and difficulties in comparing suicide statistics. The reliability of suicide statistics is influenced by whether suicide is ascertained by legal officials as in the United Kingdom and Ireland, or by medical examiners as elsewhere in Europe. In general, suicides tend to be undercounted, whereas non-suicidal deaths are very rarely misidentified as suicides. Most misclassified suicides fall into the category of undetermined deaths and are more like suicides than accidents. Underestimation is reasoned to be less than 10 per cent in the more developed countries, which allows rate comparisons between countries and over time. Despite problems in the recording of suicide, reports on suicide rates among different cultures or people suggest a true variation in suicide mortality.[1,2]

The suicide process and the act of suicide

Suicide is a mode of death usually consequent to a complex and multifaceted behaviour pattern. It is typically seen as the fatal outcome of a long-term process shaped by a number of interacting cultural, social, situational, psychological, and biological factors. Suicide is a rare, shocking, and very individual final act, which often leaves the survivors helpless. The suicide process model is used to organize and clarify the complexity of factors associated with suicide (Fig. 4.15.1.1).

Suicide is usually preceded by years of suicidal behaviour or feelings, and plans and warnings. In about half of all suicides, a previous attempt is found in the person's history, which offers, in theory, an opportunity for suicide intervention wherever suicide attempts occur. Male suicide attempts are more violent and the first attempt more likely to end in death. Successful suicide prevention calls for sensitive understanding of suicidal intent and active early intervention.[3]

Various risk or protective factors underlie suicidal behaviour. An appearance of suicidality means either an intensified effect of risk factors or a weakened effect of protective factors. For example, a separation from someone close may precipitate a suicidal imbalance in a vulnerable person due to the adverse life event as a stressor and the broken social network as a loss of social support.

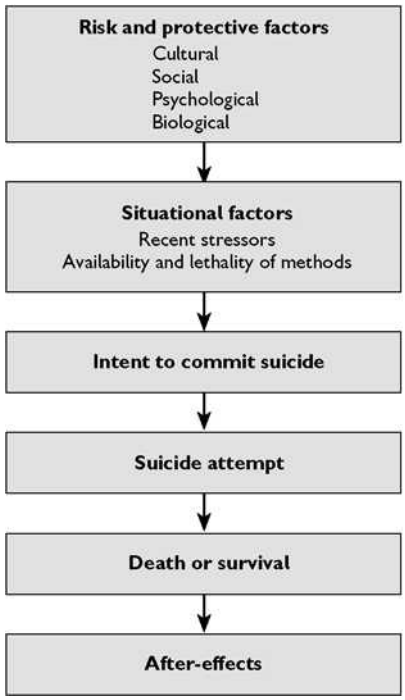


Fig. 4.15.1.1 The suicide process model.

The treating personnel and relatives of the suicide victim tend to overemphasize the meaning of the most recent events in the course of the suicide process. A precipitating factor may well be decisive in explaining the precise timing of suicide in the long course of a person's suicide process. Often, however, it also allows a simple and rational explanation in the face of the complexity of suicide.

The choice of a specific method takes place at the very end of the suicide process and represents the last possibility to intervene. Hanging is universally available and it is the most common suicide method globally. In many countries the ready access to firearms makes them potentially dangerous, especially among male adolescents and young adults. Restricting access to handguns might be expected to reduce the suicide rate of young people. Previously, domestic gas was frequently used as a suicide method, and detoxification resulted in a significant decrease in suicide rates. Nowadays, the increasing suicide rate in many Asian societies has been largely linked with pesticides and other poisons. Restrictive availability of lethal measures may also be important in the clinical treatment of individual suicidal patients. Restriction in availability of dangerous means is a strategy based on the fact that suicidal crises are often brief, suicidal acts are often impulsive, and the long-term suicide rate of serious suicide attempts is remarkably low.[4–6]

Firearms, carbon monoxide, and hanging are active suicide methods with the highest potential to cause death. Jumping from a height or leaping in the front of a moving vehicle are more passive ways, but are also highly damaging in nature. Poisoning, drowning, or wrist cutting are typically methods which leave more time for help seeking and intervention.

Imitation means learning the use of a specific suicide method from a model which is overtly available in a culture, community, institution, or mass media. Imitation may have a significant effect on the choice of a suicide method, especially at schools, in psychiatric hospital wards, and in the general population of young people. The most famous example of imitation is the effect of Goethe's novel *The Sorrows of Young Werther*, which was widely read in Europe about 200 years ago. The suicide of the hero was imitated to such an extent that authorities in several European countries banned the novel. The 'Werther effect' also appeared after the death of Marilyn Monroe; the suicide rate rose about 10 per cent over the next 10 days. Recommendations for reporting of suicide have encouraged avoidance of repetitive and excessive reports, descriptions of technical details, simplified explanations for suicide, presenting suicide as means of coping with personal problems, or glorifying suicide victims.[7]

Most suicides are solitary and private, but a few result from a pact between people to die together. Suicide pacts are exceptional, accounting for less than 1 per cent of all suicides.

Suicide always has a major impact on the survivors. Suicide is a threatening event not only among close family members, but also in the surrounding population, including treating personnel and the people at the victim's workplace. The major challenges after a suicide, in addition to a normal mourning process, are dealing with shame and guilt feelings, and the crisis of survivors. Sharing of the traumatic experience and social support should be arranged immediately and continued, if necessary, at least some months after the suicide.

Epidemiology and public health aspects of suicide

Every year one million people commit suicide, accounting for 1 to 2 per cent of total global mortality. Suicide is a leading cause of premature death, especially among young adults. It is the fifth highest cause of years of life lost in the developed world. In many westernized countries, suicide is a more frequent cause of death than traffic accidents. According to World Health Organization (**WHO**) statistics, the annual world-wide incidence of completed suicide was 16 per 100 000 persons in 2000. This means that globally one person commits suicide every minute. Suicide is estimated to represent two per cent of the total global burden of disease.

The long-term trend in suicide mortality has been increasing at least during the last 50 years. The rank order of suicide mortality in the European region in 2001 to 2003 shows that most of the countries with high suicide mortality are located in Eastern Europe (Table 4.15.1.1). Outside this region, suicide mortality has been high in Japan and China. Everywhere, the male suicide rate is clearly higher than the female rate; China is the only exception with a very high female suicide rate.

The suicide rate of elderly people has been higher than in the younger age groups almost universally. However, in many Western countries, the suicide rate for people aged 65 years and over has been declining for decades. This change is associated with the growth of the general well being and the better social and health services.

Traditionally the incidence of suicide has been low in the younger age group (15–24), but during the past 40 years the suicide rate has been rising in many Western countries, especially among young males.

Table 4.15.1.1 Suicide rates per 100 000 by country, 2001–2003

European legion	Males	Females
Lithuania	74.3	13.9
Russian Federation	69.3	11.9
Belarus	63.3	10.3
Kazakhstan	50.2	8.8
Estonia	47.7	9.8
Ukraine	46.7	8.4
Latvia	45.0	9.7
Hungary	44.9	12.0
Finland	31.9	9.8
Republic of Moldova	30.6	4.8
Czech Republic	27.5	6.8
Austria	27.1	9.3
France	26.6	9.1
Poland	26.6	5.0
Switzerland	26.5	10.6
Romania	23.9	4.7
Slovakia	23.6	3.6
Ireland	21.4	4.1
Bulgaria	21.0	7.3
Germany	20.4	7.0
Iceland	19.6	5.6
Sweden	18.9	8.1
Portugal	18.9	4.9
Luxembourg	18.5	3.5
Norway	16.1	5.8
Netherlands	12.7	5.9
Spain	12.6	3.9
Italy	11.1	3.3
United Kingdom	10.8	3.1
TFYR Macedonia	9.5	4.0
Uzbekistan	9.3	3.1
Malta	8.6	1.5
Albania	4.7	3.3
Greece	4.7	1.2
Georgia	3.4	1.1
Armenia	3.2	0.5
Azerbaijan	1.8	0.5
Other		
Japan	35.2	12.8
Republic of Korea	24.7	11.2
China (Hong Kong SAR)	20.7	10.2
Australia	20.1	5.3
Canada	18.7	5.2
United States of America	17.6	4.1
Thailand	12.0	3.8
Singapore	11.4	7.6
Kuwait	2.5	1.4

A long list of major public health concerns in the field of suicidology has emerged:

- suicidal ideation and suicide attempts are surprisingly common in the general population
- the high rate of suicides among adolescents and young adults
- unemployment as a major risk factor for suicide
- easy access to lethal suicide methods such as psychotropic or analgesic drugs, guns, and motor vehicles
- high alcohol consumption and increasing substance misuse
- undertreatment of major psychiatric disorders such as depression and schizophrenia
- suicide models projected by the mass media.

These findings indicate that rapid growth and continuous changes in society are simultaneously causing instability and disturbing the development of integration. Some regions and groups of people are inevitably affected negatively by this development, and large numbers of people are thus moving towards a greater risk of suicide.

Determinants of suicide

Usually, suicide has no single cause. It is the endpoint of an individual process, in which several interacting determinants or risk factors can be identified (Tables 4.15.1.2 and 4.15.1.3). Risk factors are by their nature cultural, social, situational, psychological, biological, and even genetic.[8]

(a) Cultural factors in suicide

Culture defines basic attitudes towards life and death, and also towards suicide in society. We still have stigma against suicide. A hundred years ago, suicide was illegal in many European countries. Similarly, most churches overtly opposed suicide and allowed suicide victims to be buried only outside the cemetery. Religion was also a major integrating force between individuals and the community. In a modern secularized society, religion is still a

Table 4.15.1.2 Risk factors for suicide: sociodemographic variables

Gender	Male
Age	Elderly
Social status	Low
Educational status	Low
Marital status	Unmarried, separated, divorced, widowed
Residential status	Living alone
Employment status	Unemployed, retired, insecure employment
Economic status	Weak (males)
Profession	Farmer, female doctor, student, sailor
Special subpopulations	Students, prisoners, immigrants, refugees, religious sects
Special institutions	Hospitals, prisons, army
Region	Uneven distribution locally by urban–rural, residential, or subcultural area
Season and time	Spring and autumn, weekend, evening, anniversary
Life events	Adverse life events such as losses and separations, criminal charges
Social support	Low
Social integration	Lacking

Table 4.15.1.3 Clinical determinants of suicide

Family history	Suicidal behaviour, mental disorders
Mental disorders	Any disorder, depression, substance use disorders, personality disorders, schizophrenia
Contact with psychiatric services	Any contacts, recent contacts, post-discharge period, psychotropic drugs
Psychiatric symptoms	Hopeless, helpless, depressive, psychotic, delirious, anxious, aggressive, introversive
Suicidal behaviour	Previous suicide attempts, suicidal ideations, death wishes, indirect gestures
Physical health	Severe physical illness such as cancer, AIDS, stroke, and epilepsy; permanent sickness
Availability of suicide methods	Easy access to lethal methods

meaningful and protective factor for many individuals in a suicidal crisis. Western culture has had a tendency to emphasize the individuals's free will and the shouldering of responsibility for one's life, while egoistic and anomic trends in society have intensified and altruism has almost disappeared. Such changes may have increased the incidence of suicide in society. The cultural background of suicide is a deep structure inherited over generations. Cultural factors also prevent rapid changes in suicidal behaviour, which is evident among immigrants, whose mode and rate of suicide usually lie somewhere between the original and the host cultures.[9]

(b) Sociological theories on suicide

Classic sociology views suicide as a social, not an individual, phenomenon. The suicide victim's moral predisposition to commit suicide, not his or her individual experiences, is felt to be the crucial factor. Moral predisposition means the degree to which the victim is involved in more or less integrated groups and in the values of those groups. Suicides are seen as a disturbance or symptom of a relationship between society and individuals. In 1897, Durkheim published his famous work on suicide and described four basic types. Anomic suicide reflects a situation where an individual is no longer guided by the society due to its weakness, like the suicide of an unemployed and rejected alcoholic without any support from society. Altruistic suicide is illustrated by a society which can exert a strong influence on an individual's decision to sacrifice his or her life, as did the captain of the *Titanic*, for example. Egoistic suicide is an individualistic decision of a person no longer dependent on others' control or opinion such as a person who has arranged an assisted suicide. Fatalistic suicide is seen as a result of strict rules in a society which have proved decisive for the destiny of an individual, for example the suicide of a person held as a slave.[10] There are also newer social theories of suicide which stress more the joint effects of social factors. The concept of social isolation has been clinically useful in understanding the socio-ecological and social-psychiatric background of suicide.[11] Some sociologists have underlined the individual meanings associated with suicide.

(c) Life events and social support

The life situation preceding suicide is typically characterized by an excess of adverse life events and recent stressors. Usually, the sum effect of events is overwhelming and more important than a single life event. Job problems, family discord, somatic illness, financial trouble, unemployment, separation, and death and illness in the family are the most common life events preceding suicide. Somatic illness and retirement are age-specifically connected with the suicides of elderly men, while separation, financial troubles, job problems, and unemployment are more common among younger men. Severely disabling somatic illness is a very important risk factor for suicide in elderly male patients. In general, suicide among men is more often related to recent stressors than it is among women.

In most cases, life events are not accidental, but are usually also dependent on the individual's own behaviour. Personality features, even mental disorders, often explain the difficulties the victim has had. Among male alcoholics, life stress is connected with family discord and separations in all age groups. Other sources of stress in alcoholic male suicides are unemployment and financial troubles, whereas in depressive non-alcoholic male victims life stress is associated more with somatic illness. Among alcoholic males, adverse life events and living alone clearly have an enhancing effect on suicidality. Among females, life events as psychosocial stress are less strongly connected with suicide. Depression and adverse interpersonal life events are more frequent contributors to female than male suicides.[12]

(d) Psychology of suicide

Early psychological theories of suicide focussed on the concept of the self. A classical example is Freud's theory assuming that self-destructive behaviour in depression represents aggression directed against a part of the self that has incorporated a loss or rejection of a love object. In his later theory of suicide, Freud presented the construction of the dual instincts, where Eros is a life-sustaining and life-enhancing drive in constant interaction with Thanatos, the aggressive death instinct.

Later psychodynamic thinking on suicide focussed more on the self in relation to others. Failures in the developmental and adaptational processes are reflected in negative self-images and distorted cognitive schemas, leading to such feelings as depression, hopelessness, rage, shame, guilt, and anxiety. It is widely held that psychological pain is found as a common element at the core of all suicidal behaviours; suicide occurs when the individual can no longer endure the pain. Most recent psychological theories of suicide accept a multifactorial causation of suicide resulting from an interaction of predisposing and precipitating factors.[13] A person moves towards a suicidal crisis depending on the stressors and presence or absence of protective factors in his or her life.

(e) Neurobiological determinants of suicide

Suicidal behaviour is highly familial. Relatives of patients who commit suicide are themselves more likely to commit suicide than relatives of patients who do not commit suicide. Liability to suicidal behaviour may be a familially transmitted trait which is partly independent on the specific mental disorders. Since the heritability of liability to suicidal behaviour appears to be on the order of 30–50 per cent, interactions with environmental factors must also be significant.

Results of adoption studies suggest that genetic factors rather than familial environmental factors are the determinants of familial concordance for suicidal behaviour. Among biological relatives

of adopted suicide victims there is a higher incidence of suicide than among the relatives of non-suicide controls or among the adoptive relatives of the suicide victims. Also identical twins have a higher concordance for suicide, attempted suicide, and suicide ideation compared with non-identical twins.[14]

The findings that genetic factors contribute to suicidal behaviour has stimulated studies aimed at identifying susceptibility genes. So far molecular genetic studies have focussed on the serontonergic pathway. Two genes have emerged as being involved in the vulnerability to suicidality: the tryptophan hydroxylase 1 (TPH1) gene, as a quantitative risk factor for suicidal behaviour, and serotonin transporter gene (5-HTTLPR), which is consistently associated with impulsive-aggressive personality traits.[15] Patients who have seriously attempted suicide by violent means have low levels of the serotonin metabolite 5-hydroxyindole acetic acid in their cerebrospinal fluid.[16] These and other neurobiological changes are discussed in Chapter 4.15.3.

Basic characteristics of the suicide victim

Persons at greatest risk of suicide are usually middle-aged or older, non-married men with poor social and economic position, and a family history of mental disorders and suicidal behaviour. Usually they are living alone, and often unemployed or with insecure employment. They also typically have marked recent life stress and a weak social network. Most suffer from depression, and feel hopeless and many have a comorbid substance abuse or personality disorder. Almost all elderly victims have comorbid physical illness or are permanently disabled. Most have previously contacted health care, and communicated their suicidal tendencies at least indirectly, although usually without receiving adequate psychiatric treatment. Half of them have previously attempted suicide.

Mental disorders and suicide

Virtually all mental disorders carry an increased risk of suicide.[17] The suicide risk in functional mental disorders is double that in substance use disorders, which in turn carry double the risk of suicide compared to organic disorders. The greatest risk of suicide among all clinical states is in attempted suicide, which carries about 40 times the expected value (Table 4.15.1.4). In anorexia nervosa and major depression, the risk is about 20-fold, and in other mood disorders and psychoses about 10 to 15 times higher than expected. In anxiety, personality, and substance use disorders the suicide risk is at lower levels, but about 5 to 10 times higher than the expected value. In subtance disorders the risk is dependent on the type of disorder, being clearly lowest in alcohol, cannabis, and nicotine abusers.[17]

Psychological autopsy studies have been used to construct an overall view of suicide by collecting all available relevant information on the victim's life preceding his or her death. In psychological autopsy studies, mental disorders of suicide victims have been assessed using DSM-diagnoses and large unselected samples. In two recent meta-analyses[18,19] the victim received at least one diagnosis on Axis I in 87 to 90 per cent of the suicides. In all studies, depressive disorders (43 per cent) and substance use disorders (26 per cent), personality disorders (16 per cent) and psychoses (9 per cent) were frequent and comorbidity was common.

Table 4.15.1.4 Rank order of suicide in mental disorders

Suicide attempt
Anorexia nervosa
Major depression
Mood disorders not otherwise specified
Reactive psychoses
Bipolar disorder
Dysthymia
Schizophrenia
Anxiety disorders
Personality disorders
Substance use disorders

In two European psychological autopsy studies[20,21] from Finland and Northern Ireland, the distribution of the principal diagnoses was similar (Table 4.15.1.5). The most common psychiatric diagnoses in suicide were major depression and alcohol dependence. Major mood disorders together comprised 42 to 36 per cent and substance use disorders, 19 to 30 per cent of all suicides. Comorbidity was a major finding in both samples, most commonly substance use disorder with major depression. A recent European psychological autopsy study gave similar results underlining, however, the role of personality disorders as a risk factor for suicide.[22]

Table 4.15.1.5 Principal diagnoses of suicide victims in Finland and Northern Ireland

Diagnosis	Finland (%) (n = 229)	Northern Ireland (%) (n = 118)
Major depression	30	31
Depressive disorder not otherwise specified	9	—
Dysthymia	—	1
Bipolar disorder	3	4
Alcohol dependence	17	24
Alcohol misuse	2	4
Other substance use disorders	—	2
Schizophrenia	7	6
Schizoaffective disorder	3	1
Other psychoses	3	4
Anxiety disorders	1	5
Adjustment disorder	3	3
Organic mental disorders	2	1
Other Axis I disorders	2	2
Personality disorder	9	3
No diagnosis	2	10
Insufficient information for assessment	7	—

The mortality risk for suicide in major depression is 20 times that expected, and 15-to 20-fold in all affective disorders. Every sixth death among depressive people treated as psychiatric patients is by suicide.[23] The risk of suicide varies across the subclasses of depression, and is related to the selection of suicidal patients for the various types of treatment. The risk is highest for depressive inpatients, even during the postdischarge period, and much lower among psychiatric outpatients, although clearly lowest for those treated for depression in primary care.[24] A meta-analysis found a hierarchy of life-time suicide prevalence: eight per cent in people ever admitted for suicidality, four per cent in patients admitted with affective disorder but not for suicidality, and two per cent in mixed inpatient and outpatient populations.[24]

Depression of suicide victims differ qualitatively from that of living controls; it seems to be more severe and accompanied more often by insomnia, weight or appetite loss, feelings of worthlessness, inappropriate guilt, and thoughts of death or suicidal ideation.[26] In addition, impulsive and aggressive behaviour, alcohol and drug abuse and dependence, and cluster B personality disorders increase the risk of suicide in individuals with major depression.[27] Inadequate and inefficient antidepressant treatment of depressed suicide victims has been a persistent finding in several studies. Less than half of suicide victims with major depression have been in contact with psychiatric care at the time of suicide. However, there is some evidence that good monitoring and maintenance treatment in high-risk groups of patients may be able to decrease their suicide rates.[28]

Alcohol and drugs, often combined, are a major risk or a precipitating factor for suicide. They may intensify the suicidal intent, offer a constantly available suicide method, worsen the somatic status of the victim, and increase the risk of complications after the attempt. Alcohol and drugs impair judgement and lower the threshold to suicide. Alcohol is detected in about every third case at the moment of suicide.[29] The lifetime risk of suicide has been estimated at 7 per cent for alcohol dependence, with only slight variation over the life.[30] The suicide rate in heavy drinking is 3.5 times and in alcohol use disorders ten times higher than that in the general population.[31] In drug dependence or abuse it is 15 times higher than expected.[17,31] The role of substance use disorders varies greatly by country. In a recent study from Estonia, alcohol dependence was found in a half of suicide victims.[32]

The suicide risk in schizophrenia appears to be almost 10 times higher than in the general population.[17] The lifetime risk of suicide in schizophrenia is estimated to be 5 per cent.[30,33] The great majority of schizophrenic patients commit suicide in the active phase of the disorder after having suffered depressive symptoms. Suicide in schizophrenia is thus less of a surprise; it is typically preceded by a previous attempt, and suicidal intent has been communicated at least as often as in non-schizophrenic suicides.[34] Schizophrenic suicide victims differ from other schizophrenic patients by having suicidal thoughts and previous suicide attempts, being more depressive, and having more positive symptoms.[35] Undertreatment, comorbidity, treatment non-compliance, and a high frequency of non-responders are also common problems among schizophrenic suicide victims. Adequacy of comprehensive care is crucial for suicide prevention in schizophrenia, especially among actively psychotic patients with recent suicidal behaviour and depressive symptoms.[36]

Personality disorders are tightly connected with suicide. Most of the suicide victims with personality disorder, especially with borderline and emotionally unstable personality disorder, have also comorbid depressive disorder or substance abuse. They often suffer from impulsive and aggressive behaviour.[20,22,27,38] This kind of comorbidity is very frequent among the young suicide victims.

Mental disorders, particularly depressive disorders, substance abuse, and antisocial behaviour have an important role in the adolescent suicides. The diagnostic distribution of mental disorders among them is surprisingly similar to that of the young and even middle-aged adults.[39]

References

1. Lönnqvist, J. (1977). Suicide in Helsinki: an epidemiological and socialpsychiatric study of suicides in Helsinki in 1960–61 and 1970–71. *Monographs of Psychiatria Fennica*, No. 8.
2. Sainsbury, P. and Jenkins, J.S. (1982). The accuracy of officially reported suicide statistics for purposes of epidemiological research. *Journal of Epidemiology and Community Health*, **36**, 43–8.
3. Isometsä, E.T. and Lönnqvist, J.K. (1998). Suicide attempts preceding completed suicide. *British Journal of Psychiatry*, **173**, 531–6.
4. Ohberg, A., Lonnqvist, J., Sarna, S., *et al.* (1995). Trends and availability of suicide methods in Finland: proposals for restrictive measures. *British Journal of Psychiatry*, **166**, 35–43.
5. Kreitman, N. (1976). The coal gas history: UK suicide rates 1960–1971. *British Journal of Preventive and Social Medicine*, **30**, 83–90..
6. Lewis, G., Hawton, K., and Jones, P. (1997). Strategies for preventing suicide. *British Journal of Psychiatry*, **171**, 351–4.
7. Hawton, K., Fagg, J., Simkin, S., *et al.* (1998). Methods used for suicide by farmers in England and Wales. *British Journal of Psychiatry*, **173**, 320–4.
8. Maris, R.W. (2002). Suicide. *Lancet*, **360**, 319-26.
9. Neeleman, J., Halpern, D., Leon, D., *et al.* (1997). Tolerance of suicide, religion and suicide rates: an ecological and individual study in 19 Western countries. *Psychological Medicine*, **27**, 1165–71.
10. Durkheim, E. (1897). *Le suicide, etude de sociologie*. Alcan, Paris. (English translation: *Suicide*. Routledge and Kegan Paul, London, 1952.)
11. Maris, R.W. (1981). *Pathways to suicide*. John Hopkins Univesrity Press, Baltimore, MD.
12. Heikkinen, M.E., Isometsä, E.T., Marttunen, M.J., *et al.* (1995). Social factors in suicide. *British Journal of Psychiatry*, **167**, 747–53.
13. Farberow, N.L. (1997). The psychology of suicide: past and present. In *Suicide: biopsychosocial approaches* (ed. A.J. Botsis, C.R. Soldatos, and C.N. Stefanis), pp. 147–63. Elsevier, Amsterdam.
14. Brent, D.A., and Mann, J.J. (2005). Family genetic studies, suicide, and suicidal behaviour. *American Journal of Medical Genetics (Semin. Med. Genet.)*, **133**(1), 13–24.
15. Bondy B., Buettner A., and Zill P. (2006). Genetics of suicide. *Molecular Psychiatry*, **11**, 336–351.
16. Mann, J.J. (2003). Neurobiology of suicidal behaviour. *Nature Reviews Neuroscience*, **4**(10), 819–828.
17. Harris, E.C. and Barraclough, B. (1997). Suicide as an outcome for mental disorders. A meta-analysis. *British Journal of Psychiatry*, **170**, 205–28.
18. Cavanagh, J.T.O., Carson, A.J., Sharpe, M. *et al* (2003). Psychological autopsy studies of suicide: a systematic review. *Psychological Medicine*, **33**, 395–405.
19. Arsenault-Lapierre, G., Kim C., and Turecki G. (2004). Psychiatric diagnoses in 3275 suicides: a meta-analysis. *BMC Psychiatry*, **4**, 37.

20. Henriksson, M.H., Aro, H.A., Marttunen, M.J., *et al.* (1993). Mental disorders and comorbidity in suicide. *American Journal of Psychiatry*, **150**, 935–40.
21. Foster, T., Gillespie, K., and McClelland, R. (1997). Mental disorders and suicide in Northern Ireland. *British Journal of Psychiatry*, **170**, 447–52.
22. Schneider, B., Wetterling, T., Sargk, D., *et al.* (2006). Axis I disorders and personality disorders as risk factors for suicide. *European Archives of Psychiatry and Clinical Neuroscience*, **256**, 17–27.
23. Wulsin, L.R., Vaillant, G.E., and Wells, V.E. (1999). A systematic review of the mortality of depression. *Psychosomatic Medicine*, **61**, 6–17.
24. Simon, G.E. and VonKorff, M. (1998). Suicide mortality among patients treated for depression in an insured population. *American Journal of Epidemiology*, **147**, 155–60.
25. Bostwick, J.M. and Pankratz, V.S. (2000). Affective disorders and suicide risk: a reexamination. *American Journal of Psychiatry*, **157**, 1925–32.
26. McGirr, A., Renaud, J., Seguin, M., *et al.* (2007). An examination of DSM-IV depressive symptoms and risk for suicide completion in major depressive disorder: A psychological autopsy study. *Journal of Affective Disorders*, **97**, 203–9.
27. Dumais, A., Lesage, A.D, Alda, M., *et al.* (2005). Risk factors for suicide completion in major depression: a case-control study of impulsive and aggressive behaviors in men. *American Journal of Psychiatry*, **162**, 2116–24.
28. Isometsä, E.T., Henriksson, M.M., Aro, H.M., *et al.* (1994). Suicide in major depression. *American Journal of Psychiatry*, **151**, 530–6.
29. Ohberg, A., Vuori, E., Ojanperä, I., *et al.* (1996). Alcohol and drugs in suicides. *British Journal of Psychiatry*, **169**, 75–80.
30. Inskip, H.M., Harris, E.C., and Barraclough, B. (1998). Lifetime risk of suicide for affective disorder, alcoholism and schizophrenia. *British Journal of Psychiatry*, **172**, 35–7.
31. Willcox, H.C., Conner, K.R., and Caine, E.D. (2004). Association of alcohol and drug use disorders and completed suicide: an empirical review of cohort studies. *Drug and Alcohol Dependence*, **76S**, S11-S19.
32. Kõlves, K., Värnik, A., Tooding, L-M., *et al.* (2006). The role of alcohol in suicide: a case-control psychological autopsy study. *Psychological Medicine*, **36**, 923–930.
33. Palmer, B.A., Pankratz, V.S., and Bostwick, J.M. (2005). The lifetime risk of suicide in schizophrenia. *Archives of General Psychiatry*, **62**, 247–253.
34. Heila, H., Isometsä, E., Henriksson, M.H., *et al.* (1997). Suicide and schizophrenia: a nationwide psychological autopsy study on age- and sex-specific clinical characteristics of 92 suicide victims with schizophrenia. *American Journal of Psychiatry*, **154**, 1235–42.
35. Kelly, D.L., Shim, J-C., Feldman, S.M, *et al.* (2004). Lifetime psychiatric symptoms in persons with schizophrenia who died by suicide compared to other means of death. *Journal of Psychiatric Research*, **38**, 531–6.
36. Heilä, H., Isometsä, E., Henriksson, M.H., *et al.* (1999). Suicide victims with schizophrenia in different treatment phases and adequacy of antipsychotic medication. *Journal of Clinical Psychiatry*, **60**, 200–8.
37. Zouk, H., Tousignant, M., Sequin, M., *et al.* (2006). Characterization of impulsivity in suicide completers: Clinical, behavioral and psychosocial dimensions. *Journal of Affective Disorders*, **92**, 195–204.
38. Brezo, J., Paris, J., and Turecki, G. (2006). Personality traits as correlates of suicidal ideation, suicide attempts, and suicide completions: a systematic review. *Acta Psychiatrica Scandinavica*, **113**, 180–206.
39. Marttunen, M.J., Aro, H.M., Henriksson, M.M., *et al.* (1991). Mental disorders in adolescent suicide: DSM-III-R axes I and II diagnoses in suicides among 13- to 19-year-olds in Finland. *Archives of General Psychiatry*, **48**, 834–9.

4.15.2 Deliberate self-harm: epidemiology and risk factors

Ella Arensman and Ad J. F. M. Kerkhof

Introduction

Deliberate self-harm (DSH) refers to behaviour through which people deliberately inflict acute harm upon themselves, poison themselves, or try do so, with non-fatal outcome. These behaviours are somehow linked to, but do not result in, death. Common to these behaviours is that they occur in conditions of emotional turmoil. In former days these behaviours were often regarded as failed suicides. However, this view did not appear to be correct, and the great majority of patients in fact do not try to kill themselves. Therefore, the term deliberate self-harm was introduced to describe the behaviour without implying any specific motive.[1] But this too has some disadvantages because there is a temporal association between non-fatal and fatal suicidal behaviour; many people who die by suicide have engaged in DSH before. Thus, Kreitman *et al.*[2] introduced the concept of parasuicide to describe behaviour that, mostly without the intention to kill oneself, communicates a degree of suicidal intent. However, both terms, deliberate self-harm and parasuicide, are still somewhat confusing, because in practice they include people who really have the intent of killing themselves but survive the attempt. The difficulty of finding a good terminology for these behaviours is reflected in differences in research populations in empirical studies: some studies are limited to self-poisoning only (overdose), a few studies are restricted to self-injury (wrist cutting) only, some to self-poisoning and self-injury combined, and some studies include behaviours in which, due to last-moment intervention from others, there was no actual self-harm inflicted at all. In recent years the term self-harm is being used in the United Kingdom and North America since the adjective 'deliberate' is not favoured by patients, particularly those who repeatedly engage in acts of self-harm.[3]

In this chapter, we will use the term deliberate self-harm interchangeably with attempted suicide to refer to non-fatal suicidal behaviours in which there may have been an intention to die, however ambiguous this intention may have been, and irrespective of other intentions that may have been operating at the same time. It should be stressed that in deliberate self-harm many motives may play a role simultaneously, even contradictory motives such as the hope of being rescued and the wish to continue living. Intentions may vary from attention seeking or communication of despair, appeal for help, to a means for stress reduction. Common to these behaviours is that they are motivated by change: people want to bring about changes in their present situation through the actual or intended harm or unconsciousness inflicted upon the body. Deliberate self-harm may be defined as follows.[4]

> An act with non-fatal outcome, in which an individual deliberately initiates a non-habitual behaviour that, without intervention from others, will cause self-harm, or deliberately ingests a substance in excess of the prescribed or generally recognised therapeutic dosage, and which is aimed at realising changes which the subject desired via the actual or expected physical consequences.

This definition covers deliberate non-fatal suicidal behaviours. Not included are accidental cases of self-poisoning, accidental overdoses of opiates, or self-harmful acts by persons who do not anticipate the consequences of their actions. It does not include automutilation, which is an habitual, often obsessive act of inflicting (minor) self-harm, mostly without a conscious intent of changing the present situation, as with certain persons with learning disability.

Clinical features

Deliberate self-harm can have very different motivations, varying from an intention to die to a cry for help. These behaviours may be well prepared or carried out impulsively, and may have different physical consequences. The degree of lethality and the degree of medical seriousness of the consequences thus depend upon intention, preparation, knowledge and expectations of the method chosen, and sometimes upon coincidental factors such as intervention from others.

It is often difficult to assess the true intent of DSH. Because of fear for consequences, such as admission to a psychiatric hospital, or because of psychological defence mechanisms, people sometimes deny or conceal their intention to die. They also may exaggerate their intention to die in order to receive help. Sometimes people engage in potentially highly lethal self-harming behaviour without any wish to die, for example when they do not have adequate knowledge of the medication used. People who present at a general hospital with minor self-injury or minor self-poisoning may have had strong intentions to die but had insufficient knowledge of the lethality of the method. Therefore, one cannot always reliably infer what the precise meaning of the behaviour was, either from its overt characteristics or from the person's self-report. Among a large sample of adolescents aged 15 and 16 years, Rodham *et al.* found that adolescents who took an overdose more often expressed a wish to die compared to those who engaged in self-cutting.[5] Motives associated with self-cutting were self-punishment and interruption, i.e. trying to get relief from a terrible state of mind.

Epidemiology

In the 1960s and 1970s, there was a sharp increase in the number of people treated in hospitals in Europe, the United States and Australia because of intentional overdoses or self-injury. In the 1980s several studies showed a stabilization.[6,7] In the early 1990s these numbers increased further in some regions.[8,9] The absolute number of persons treated for deliberate self-harm in general hospitals, however, does not adequately reflect the size of the problem. These numbers should be calculated against the size and the characteristics of the population in the areas that are being served by the hospitals. Furthermore, in some countries DSH patients are treated by general practitioners when there is no need for hospital admission. In many instances emergency attendance for overdosing is not even registered. Except for Ireland, where a National Registry of Deliberate Self-Harm has been established,[10] there are no national registries that reliably monitor trends in DSH treated in general hospitals. Even though DSH is considered a major problem in the United States, clinical epidemiological research into DSH is uncommon.[11] Also, few epidemiological studies on DSH originate from other parts of the world.

Changes over time

In Edinburgh and Oxford, in the United Kingdom, there has been continuous monitoring of deliberate self-harm over a long period of time, where characteristics of persons engaging in DSH have been related to the corresponding population.[7,12,13] In these two cities trends in DSH rates have been documented reliably. After a period of stabilization in the 1980s a marked increase was observed. Between 1985 and 1995 the rates of DSH in Oxford increased by 62 per cent in males and 42 per cent in females. The increase in DSH has been most marked among young males. A similar trend has been observed in North Worcestershire where hospital referred cases of DSH were monitored over a period of 20 years (1981–2000).[14]

In Canada, the DSH rate was estimated to be around 304 per 100 000.[15] In the United States National Institute of Mental Health's Epidemiological Catchment Area study (1980–1985) it was found that 2.9 per cent of the respondents had engaged in DSH at some point of time.[16]

So far only one international multicentre study into deliberate self-harm has been conducted taking into consideration the methodological pitfalls outlined above. The World Health Organization (WHO) initiated a collaborative multicentre study in 16 regions in Europe using the same methodology, definition, and case-finding criteria.[9,17] The findings were related to the size and characteristics of the corresponding general population in order to investigate rates, trends, risk factors, and social indicators. Most of the epidemiological data presented here have been drawn from that study.[9,18]

Differences between countries and regions

There is widespread variation between countries with regard to rates of deliberate self-harm. Based on the latest available data for the years 1995–1999, overall, DSH rates (person-based) were highest in the United Kingdom (Oxford), Belgium (Ghent), Hungary (Pecs) and Finland (Helsinki), with female rates per 100 000 ranging from 83 in Padova (Italy) to 433 in Oxford. Male DSH rates per 100 000 ranged from 53 in Umea (Sweden) to 337 in Oxford.[9] Looking at trends over time, an average decrease for male DSH rates of 13 per cent was found comparing average person-based rates for 1989/1993 to the period 1995/1999, with the greatest reduction (70 per cent) in Innsbruck (Austria). For female DSH rates the average decrease in the same period was lower (4 per cent), with the greatest reduction (31 per cent) in Sor-Trondelag (Norway). In addition to medically referred cases of deliberate self-harm, community-based studies show that an even higher proportion of DSH appears to be 'hidden' from health care services.[19]

Differences between catchment areas in deliberate self-harm rates in the WHO/EURO study have been studied in relation to socio-economic characteristics of these areas.[9,20] No correlations were found with most of the social and economical factors supposedly related to DSH, such as population density, urban–rural distribution, proportion working in agriculture forestry or fishery, sex ratio, percentage aged 40 and over, number of people per household, percentage people living alone, percentage single parent families, per capita income, unemployment rate, life expectancy, mortality rate, infant mortality, crimes per year per 1000, and per capita alcohol consumption. Only two characteristics of the catchment areas seemed to be related to DSH rates: the percentage

of divorced people in the area and the percentage receiving social security. Family stability and the percentage of the population relying on welfare both seem to be related to the frequency of DSH, but the interpretation of these findings is difficult because one would expect the other related social indicators of societal cohesion to covary as well.

It is important, however, to realize that the characteristics mentioned above relate to regions or countries, and do not relate to individuals. At individual level, characteristics such as unemployment play an important role, but this does not mean that unemployment rates do explain high DSH rates in a region.[21,22] This relationship holds only for some regions and not for others, as is documented repeatedly.[23] The effect of exposure to risks factors may be due to contextual effects, which arise if individuals' risks of suicidal behaviour depends not only on their personal exposure to risk or protective factors, but also on how these are distributed in their social, cultural or economic environments.[24,25] In a small area study in South East London, Neeleman *et al.* found that the DSH rate of minority groups relative to the white group was low in some areas and high in other areas.

Cultural variation in DSH has been documented from India,[26] Sri Lanka,[27] and Pakistan,[28] and from ethnic groups within Western societies, such as the Inuit in Canada.[29] Neeleman *et al.*[30] studied ethnic differences in DSH in Camberwell, London, and found considerable differences between the DSH rates for white people and for British-born Indian females and African-Caribbeans.[30] Indian females had a particularly high rate, 7.8 times that of white females. Marriage problems seem to be related to DSH in Asian countries such as India, Pakistan, Sri Lanka, and China. Young married women may have serious difficulties after moving in with their husbands' extended families. Dowry problems and problems with in-laws are thought to be precipitants of attempted suicide among young married women. In Asian countries the methods used in DSH reflect differences in accessibility. Self-poisoning with organophosphate pesticides and other household poisons is prevalent. As in the Western world DSH appears to reflect feelings of hopelessness and helplessness in adverse living conditions with no prospect of improvement. Women tend to be more powerless to bring about changes in their living conditions. In Sri Lanka, the continuous warfare, poverty, and the lack of opportunities at home and abroad frustrates the young who are relatively well educated.[27]

Sex and age

In all but one centre (Helsinki) of the WHO Multicentre Study on Suicidal Behaviour the female DSH rates were higher than the male rates. Across the participating regions, on average, the rates for females were 1.5 times higher than those for men. DSH rates were consistently higher among those in the young age groups, with the highest person-based male DSH rates in the age group 25–34 years, whereas for females in most centres the highest rates were found in the age group 15–24 years.[9]

Sociodemographic characteristics

Single and divorced people were over-represented among people who engaged in DSH in the WHO/EURO study.[9,18] Nearly half of the males and 38 per cent of the females were never married. An interaction effect was found for age in that the proportion of single persons among those engaging in deliberate self-harm reduced with increasing age, whereas the proportion of divorced, separated, and widowed people increased with age. Among deliberate self-harm patients who were economically active, a high percentage was unemployed. Based on average DSH rates over the period 1995–1999, 26 per cent of the males and 14 per cent of the females were unemployed.[9]

These findings are consistent with outcomes of earlier research conducted in the United Kingdom, where socio-economic deprivation (low social class and unemployment) repeatedly appear as characteristics of the DSH populations.[22]

These findings indicate that DSH patients disproportionately have had low education, and have high levels of unemployment, poverty, and divorce. The findings may be partly related to underlying common causes, such as the presence of psychiatric disorders, but they also suggest the influence of sociological factors impacting on a relatively economically deprived group in society with a greater share of adversity.[24] Socio-economic deprivation is a well-established determinant of psychiatric morbidity and DSH.[32,33] In contrast with completed suicide, where the presence of psychiatric disorders is well documented (up to 95 per cent of suicides may have suffered from a psychiatric disorder), psychiatric disorders are much less frequent among those who deliberately harm themselves. Among those who engage in DSH for the first time in their lives, the prevalence of psychiatric disorders may be rather low; among repeaters, psychiatric morbidity is considerable.[34,35]

Methods

Methods used in deliberate self-harm are mostly 'non-violent'. In the WHO Multicentre Study, 65 per cent of males and 82 per cent of females took an overdose, based on average DSH rates for the period 1995–1999. Cutting, mostly wrist cutting, was employed in 16 per cent of male cases and 9 per cent of female cases. There are some differences between European countries in the use of particular methods. Based on the years 1995–1999, a relatively high percentage of self-cutting was found in Tallinn (Estonia, 50 per cent), Ljubljana (Slovenia, 30 per cent), and Innsbruck (Austria, 26 per cent). In Szeged (Hungary), 19 per cent of males and 15 per cent of females used poisoning with pesticides, herbicides, or other toxic agricultural chemicals, whereas in other regions this ranged from 0–3 per cent.[9] In Sor-Trondelag, Norway, higher percentages engaged in DSH by deliberate alcohol overdose (6 per cent of males and 5 per cent of females). In general, somewhat older men used the method of jumping or jumping in front of a moving object. In the Oxford studies between 1985 and 1995, 88 per cent of all episodes involved self-poisoning, 8 per cent involved self-injury, and 4 per cent involved both. There was an increase in the use of paracetamol from 31 per cent of poisoning cases in 1985 to 50 per cent in 1995.[6] There was also an increase in antidepressant overdoses and a decrease in overdoses of minor tranquillizers and sedatives. Comparing the early 1990s with the late 1990s, a slight increase was observed in overdose by medication.[9] For all regions the methods used in DSH acts did not covary significantly with age.

The differences in methods between countries may be related to differences in the accessibility of certain methods. Until 1998 paracetamol was available in large quantities in the United Kingdom, unlike other European countries.[36,37] The ingestion of

alcohol during or before the act sometimes can be considered to be a part of the actual method of DSH (when used to bring about unconsciousness, or to increase the risk of a fatal outcome), as part of the preparation (to lower the threshold for engaging in an act of DSH, because of disinhibition), or as a long-term risk factor. Hawton *et al.*[38,39] found that 22 to 26 per cent of DSH patients had consumed alcohol at the time of the act (males more frequent than females), and that 44 to 50 per cent had consumed alcohol during the 6 h before the DSH acts, this again being more common in males than in females. About 28 per cent of DSH patients in Oxford appeared to be substance misusers (alcohol and drugs).

General population self-report surveys

General population epidemiological surveys of adolescents indicate that DSH acts occur more frequently than suggested by hospital statistics.[40,41] A number of surveys have been conducted to estimate the prevalence of DSH. Most of these surveys concerned adolescents and were administered anonymously. Most questionnaire studies revealed that between 1 and 20 per cent of respondents had engaged in DSH at some point in time.[41–44] However, the methodology used in the various studies varies considerably and therefore limits comparison of the outcomes. In the study by Hawton *et al.*[41] a minority (12.6 per cent) of adolescents who had engaged in DSH had presented to hospital.

Lifetime prevalence

Based upon the rates from the WHO/EURO study the lifetime prevalence of deliberate self-harm should be around 3 per cent for females and 2 per cent for males, with some variations between countries and regions.[9] DSH acts that did not lead to medical treatment at a hospital or general practitioner's surgery are very difficult to study, because of the limited validity of self-report data. Based on two large community-based surveys, the lifetime prevalence of suicidal ideation varied from 2.6 to 25.4 per cent and for deliberate self-harm this varied from 0.4 to 4.2 per cent.[45,46]

Classification

As previously mentioned, there is a considerable variety of behaviours within the broad category of non-fatal suicidal behaviour. A review of classification studies[34] revealed three types of DSH patients: a 'mild' type, a 'severe' type, and a 'mixed' type in between.

The mild type of DSH encompasses mostly relatively non-violent methods followed by non-serious physical injury. Young age, living together, few precautions to prevent discovery, low level of suicidal preoccupation, low suicidal intent, interpersonal motivation are all characteristics associated with mild forms of attempted suicide/deliberate self-harm. The severe category consists mostly of relatively hard methods followed by serious physical consequences. Older age (over 40), many precautions to prevent discovery, high level of suicidal preoccupation, high suicidal intent, self-directed motivation, often relocated, previous attempted suicides, depression, drug dependence, a high degree of overall dysfunctioning, poor physical health, and previous psychiatric treatment are all characteristics associated with the concept of 'severe' deliberate self-harm. The risk of repetition is greater in the severe type. In between, in the mixed type of DSH, the DSH acts and patients involved show mixed characteristics, which makes this type harder to identify in medical practice.

In order to further refine the classification of deliberate self-harm, Arensman[47] included psychological and personal history variables and these characteristics were studied in relation to recurrent DSH in a follow-up period of 1 year. **The mild DSH type** was validated, approximately 40 per cent of the total sample, as being predominantly younger than 30 years of age, single, living alone or with parents, and having minor injuries because of the index attempt. The mean number of previous DSH acts was 3.7. The repetition rate in the follow-up period for this group was 27 per cent. In the older age group, two groups were distinguished: a moderate group and a group with an extremely high risk for non-fatal repetition were identified. The high-risk group, consisting of approximately 28 per cent of the total sample, suffered more physical injury as a consequence of their deliberate self-harm.

The high-risk group consisted predominantly of females in the age group 30 to 49 years who were divorced or separated, living alone, and who were economically inactive. Most of them had engaged in previous DSH acts (mean number: 5). They had histories of traumatic life events that mostly started early in life. The high-risk group showed the highest scores on depression, hopelessness and expression of state-anger, and two-thirds were diagnosed as having borderline personality disorder. In the follow-up at least 75 per cent engaged in repeated DSH.

The moderate group was characterized by low levels of physical injury following their DSH, they were predominantly aged at least 30 years and married, and scored intermediate on measures of depression, hopelessness, and anger. The mean number of previous attempts was 2.3, and 33 per cent made one or more repeated DSH acts in the follow-up. Surprisingly, this classification into three types of non-fatal suicidal behaviours did not correspond to the levels of reported suicide intent nor to the levels of the different motivations reported (to die, to appeal, to lose consciousness, revenge), underlining the difficulty of classifying DSH according to intentions. Rodham *et al.*[5] found evidence for different subgroups of DSH patients based on DSH methods and motives. For example, they identified a subgroup of female adolescents who engaged in self-cutting and who reported *self-punishment* as the primary motive, followed by *trying to get relief from a terrible state of mind*. Furthermore, *wish to die* as a motive was significantly more often reported by those who took an overdose compared with those who engaged in self-cutting.

Aetiology

The last psychological step towards deliberate self-harm is always set in conditions of emotional turmoil, an emotional crisis. Essential in crisis is the absence of any positive outlook towards the future. People completing suicide do not expect any improvement of their situation in the near or distant future. People who engage in deliberate self-harm indicate that their future is hopeless, but they still seem to have a faint hope, however ambiguous this may be, that the future might improve. In this way deliberate self-harm may be conceived as a self-invented form of crisis intervention. Studies into the cognitive functioning of DSH patients indeed show a global and stable form of negative anticipations and absence of positive anticipations towards the future, probably as a consequence of disturbances in their autobiographical memory, i.e. an

overgeneral memory.[48,49] Whenever these anticipations remain overgenerally negative after an act of DSH, it is likely that hopelessness will increase and that this behaviour will occur again.[50]

Precipitants

Difficulties or conflicts that may bring the person to believe that his or her future is without hope can trigger the psychological crisis resulting in deliberate self-harm. DSH is often precipitated by disharmony with key figures, work-related problems, financial difficulties, or physical illnesses. Long-standing relationship problems or feelings of loneliness are especially common. People who engage in DSH have a weak social support system,[48] and they report relational difficulties as major problems in life. They show deficits in interpersonal problem solving, and their future holds no promise. Their emotional status can best be described as a state of learned helplessness, a situation of a blocked escape, in which no solution exists for a perceived insurmountable adversity.[51] This leads to the question as to why these persons have developed such helpless attitudes.

Long-term vulnerability factors

The conflicts experienced by DSH patients in the days before an act of DSH are not different from the same conflicts they have experienced over and over again. Not only recent life events, but also the life events that occurred in their past are important.[47] Many DSH patients, males as well as females, have had traumatic childhood experiences, including physical and emotional neglect, broken homes, other unstable parental conditions, violence, sexual and physical abuse, incest, parents who had psychiatric treatment, who were alcoholics and/or addicted to opiates. Women who have been abused have a much greater probability of becoming a repeater later. In addition they often develop poor relationships, lack self-esteem, and experience overwhelming feelings of helplessness and hopelessness. Any trigger, for example an argument with a friend, may be sufficient to provoke suicidal ideation and behaviour.

DSH patients not only suffer from helplessness with regard to interpersonal conflicts, they also tend to be powerless in other domains of life. The DSH population disproportionately consists of unemployed persons, from low social classes, with low educational levels, economically deprived, divorced, disabled, addicted, incarcerated, and/or lonely. Many have received in- or outpatient psychiatric treatment. These findings are somewhat complicated by the fact that many of these vulnerability factors are strongly interrelated. Unemployment, addiction, and unstable partnership relations all may be caused by psychiatric diseases. For example, unemployment and DSH may both be a consequence of addiction. However, it is not fair to assume that most of the economic deprivation of suicidal patients is explained by their psychiatric condition. The considerable differences between nations in the prevalence of DSH support the importance of socio-economic conditions.

Course and prognosis

Repetition is one of the core characteristics of suicidal behaviour. Among those who die by suicide up to 40 per cent have a history of previous DSH acts.[52] Among DSH patients 'repeaters' are more common than 'first-evers'. Between 30 and 60 per cent of DSH patients engaged in previous acts, and between 15 and 25 per cent did so within the last year.[3,7,53,54]

Risk of suicide after deliberate self-harm

Prospectively, DSH patients have a high risk of dying by suicide. Between 10 and 15 per cent eventually die because of suicide.[11,53] The connection between DSH and suicide lies between 0.5 and 2 per cent after 1 year and above 5 per cent after 9 years.[11] Mortality by suicide is higher among DSH patients who have engaged in previous acts of DSH.[55,56] The risk of suicide after deliberate self-harm for males is nearly twice the female risk, the risk being particularly high in the first year.[57,58] Alcohol and drug abuse and related social deterioration are risk factors for subsequent suicide,[59] as are psychiatric diagnosis (affective disorders, schizophrenia, personality disorders), and a highly lethal non-impulsive index act of DSH. In a large European study focusing on young deliberate self-harm patients, positive correlations were found between rates of DSH and suicide for both males and females, with a statistically significant association among males aged 15–24.[60]

Repetition of deliberate self-harm

Risk of repeated DSH is highest during the first year after an act of DSH, and especially within the first 3 to 6 months.[55,57,61] In the WHO/EURO Multicentre Study on Suicidal Behaviour it was found that at least 54 per cent of DSH patients had engaged in a DSH act before, 30 per cent at least twice. Prospectively, 30 per cent of DSH patients made at least one repeated attempt in a 1-year follow-up.[62,63]

It is hoped for that knowledge of antecedents or risk factors may foster early identification of persons at risk, and improvement of treatment. Many studies have tried to identify risk factors or antecedents and some of these by now are well known. Sociodemographic risk factors associated with repetition are belonging to the age group of 25 to 49 years, being divorced, unemployed, and coming from low social class. Psychosocial characteristics of repeaters are substance abuse, depression, hopelessness, personality disorders, unstable living conditions/living alone, criminal records, previous psychiatric treatment, and a history of stressful traumatic life events, including broken homes and family violence, especially physical and mental maltreatment by partners. Prospectively, a history of previous attempts is one of the most powerful predictors of future non-fatal suicide attempts.[34,51,64,65]

Conclusions

Deliberate self-harm is a major problem in many contemporary societies. DSH seems to reflect the degree of powerlessness and hopelessness of young people with low education, low income, unemployment, and difficulties in coping with life stress. As such, non-fatal suicidal behaviour should be a major concern for politicians. There are substantial differences between communities in the prevalence of deliberate self-harm. This suggests that some communities better meet the needs of their underprivileged youngsters than others do, but we barely understand the differences between communities and nations. Preventive action therefore is difficult to design. There is a need for a better nationwide continuous registration of DSH and related socio-economic conditions. There is also a need for better mental health care management of DSH patients, and for experimental studies on the prevention of

repetition. Although we know that persons who engage in DSH are at high risk for future fatal and non-fatal suicidal behaviour, development of effective intervention, and prevention programmes is a key priority.

Further information

Hawton, K. (ed.) (2005). *Prevention and treatment of suicidal behaviour: from science to practice*. Oxford University Press, Oxford.

De Leo, D., Bille-Brahe, U., Kerkhof, A., and Schmidtke, A. (eds.) (2004). *Suicidal behaviour: theories and research findings*. Hogrefe and Huber, Göttingen, Germany.

Van Heeringen, K. (ed.) (2001). *Understanding suicidal behaviour. The suicidal process approach to research, treatment and prevention*. John Wiley & Sons Ltd., Chichester, UK.

Hawton, K. and Van Heeringen, K. (eds.) (2000). *The international handbook of suicide and attempted suicide*. John Wiley & Sons, Chichester, UK.

Website of the International Association for Suicide Prevention (IASP): http://www.med.uio.no/iasp/

References

1. Morgan, H.G., Barton, J., Pottle, S., *et al.* (1976). Deliberate self-harm: a follow-up study of 279 patients. *The British Journal of Psychiatry*, **128**, 361–8.
2. Kreitman, N., Philip, A.E., Greer, S., *et al.* (1969). Parasuicide. *The British Journal of Psychiatry*, **115**, 746–7.
3. Skegg, K. (2005). Self-harm. *Lancet*, **366**, 1471–83.
4. Platt, S., Bille-Brahe, U., Kerkhof, A., *et al.* (1992). Parasuicide in Europe: the WHO/EURO multicentre study on parasuicide. I. Introduction and preliminary analysis for 1989. *Acta Psychiatrica Scandinavica*, **85**, 97–104.
5. Rodham, K., Hawton, K., and Evans, E. (2004). Reasons for deliberate self-harm: comparison of self-poisoners and self-cutters in a community sample of adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**, 80–7.
6. Hawton, K. and Fagg, J. (1992). Trends in deliberate self-poisoning and self-injury in Oxford, 1976–1990. *British Medical Journal*, **304**, 1409–11.
7. Platt, S., Hawton, K., Kreitman, N., *et al.* (1988). Recent clinical and epidemiological trends in parasuicide in Edinburgh and Oxford: a tale of two cities. *Psychological Medicine*, **18**, 405–18.
8. Schmidtke, A., Bille Brahe, U., De Leo, D., *et al.* (1996). Attempted suicide in Europe: rates, trends and sociodemographic characteristics of suicide attempters during the period 1989–1992. Results of the WHO/EURO multicentre study on parasuicide. *Acta Psychiatrica Scandinavica*, **93**, 327–38.
9. Schmidtke, A., Bille-Brahe, U., De Leo, D., *et al.* (2004). *Suicidal behaviour in Europe: results from the WHO/EURO multicentre study on suicidal behaviour*. Hogrefe & Huber, Göttingen, Germany.
10. National Suicide Research Foundation. (2007). *National registry of deliberate self harm report—2006*. National Suicide Research Foundation, Cork.
11. Owens, D., Horrocks, J., and House, A. (2002). Fatal and non-fatal repetition of self-harm. Systematic review. *The British Journal of Psychiatry*, **181**, 193–9.
12. Kreitman, N. (1977). *Parasuicide*. Wiley, London.
13. Hawton, K., Fagg, J., Simkin, S., *et al.* (1997). Trends in deliberate self-harm in Oxford, 1985–1995. *The British Journal of Psychiatry*, **171**, 556–60.
14. O'Loughlin, S. and Sherwood, J. (2005). A 20-year review of trends in deliberate self-harm in a British town, 1981–2000. *Social Psychiatr y and Psychiatric Epidemiology*, **40**, 446–53.
15. Sakinofsky, I. (1996). The epidemiology of suicide in Canada. In *Suicide in Canada* (eds. A.A. Leenaars, S. Wenckstern, I. Sakinovsky, D. Dyck, M. Kral, and R. Bland). University of Toronto Press, Toronto.
16. Moscicki, E.K., O'Carroll, P., Rae, D.S., *et al.* (1988). Suicide attempts in the epidemiologic catchment area study. *The Yale Journal of Biology and Medicine*, **61**, 259–68.
17. Kerkhof, A.J.J.M., Schmidtke, A., Bille Brahe, U., *et al.* (1994). *Attempted suicide in Europe*. DSWO-Press/World Health Organization, Leiden and Copenhagen.
18. Schmidtke, A., Bille Brahe, U., De Leo, D., *et al.* (1996). Attempted suicide in Europe: rates, trends and sociodemographic characteristics of suicide attempters during the period 1989–1992. Results of the WHO/EURO multicentre study on parasuicide. *Acta Psychiatrica Scandinavica*, **93**, 327–38.
19. Evans, E., Hawton, K., and Rodham, K. (2004). Factors associated with suicidal phenomena in adolescents: a systematic review of population-based studies. *Clinical Psychology Review*, **24**, 957–79.
20. Bille-Brahe, U., Andersen, K., Wasserman, D., *et al.* (1996). The WHO-EURO multicentre study: risk of parasuicide and the comparability of the areas under study. *Crisis*, **17**, 32–42.
21. Platt, S. (1984). Unemployment and suicidal behaviour: a review of the literature. *Social Science & Medicine*, **19**, 93–115.
22. Platt, S. and Dyer, J. (1987). Psychological correlates of unemployment among male parasuicides in Edinburgh. *The British Journal of Psychiatry*, **151**, 27–32.
23. Adam, K.S. (1990). Environmental, psychosocial, and psychoanalytic aspects of suicidal behavior. In *Suicide over the life cycle* (eds. S.J. Blumenthal and D.J. Kupfer), pp. 39–96. American Psychiatric Press, Washington, DC.
24. Neeleman, J., Wilson-Jones, C., and Wessely, S. (2001). Ethnic density and deliberate self harm: a small area study in south east London. *Journal of Epidemiology and Community Health*, **55**, 85–90.
25. Neeleman, J. (2002). Beyond risk theory: suicidal behaviour in its social and epidemiological context. *Crisis*, **23**, 114–20.
26. Latha, K.S., Bhat, S.M., and D'Souza, P. (1996). Suicide attempters in a general hospital unit in India: their socio-demographic and clinical profile—emphasis on cross-cultural aspects. *Acta Psychiatrica Scandinavica*, **94**, 26–30.
27. Eddleston, M., Rezvi Sheriff, M.H., and Hawton, K. (1998). Deliberate self harm in Sri Lanka: an overlooked tragedy in the developing world. *British Medical Journal*, **317**, 133–5.
28. Khan, M.M., Islam, S., and Kundi, A.K. (1996). Parasuicide in Pakistan: experience at a university hospital. *Acta Psychiatrica Scandinavica*, **94**, 264–7.
29. Kirmayer, L.J., Malus, M., and Boothroyd, L.J. (1996). Suicide attempts among Inuit youth: a community survey of prevalence and risk factors. *Acta Psychiatrica Scandinavica*, **94**, 8–17.
30. Neeleman, J., Jones, P., van Os, J., *et al.* (1996). Parasuicide in Camberwell—ethnic differences. *Social Psychiatry and Psychiatric Epidemiology*, **31**, 284–7.
31. Hawton, K. and Catalan, J. (1982). *Attempted suicide. A practical guide to its nature and management*. Oxford University Press, Oxford.
32. Gunnell, D.J., Peters, T.J., Kammerling, R.M., *et al.* (1995). Relation between parasuicide, suicide, psychiatric admissions, and socioeconomic deprivation. *British Medical Journal*, **311**, 226–30.
33. Congdon, P. (1996). Suicide and parasuicide in London; a small-area study. *Urban Studies*, **1**, 137–58.
34. Arensman, E. and Kerkhof, A.J.F.M. (1996). Classification of attempted suicide: a review of empirical studies, 1963–1993. *Suicide & Life-threatening Behavior*, **26**, 46–67.
35. De Leo, D., Bille-Brahe, U., Kerkhof, A.J.F.M., *et al.* (2004). *Suicidal behaviour: theories and findings*. Hogrefe & Huber, Göttingen, Germany.

36. Hawton, K., Simkin, S., Deeks, J., *et al.* (2004). UK legislation on analgesic packs: before and after study of long-term effect on poisonings. *British Medical Journal*, **329**, 1076.
37. Gunnell, D., Hawton, K., Murray, V., *et al.* (1997). Use of paracetamol for suicide and non-fatal poisoning in the UK and France: are restrictions on availability justified? *Journal of Epidemiology and Community Health*, **51**, 175–9.
38. Hawton, K., Fagg, J., Simkin, S., *et al.* (1997, 1998). *Deliberate self-harm in Oxford, 1996, 1997. Reports from the Oxford monitoring system for attempted suicide*. Warneford Hospital, Oxford.
39. Hawton, K., Simkin, S., and Fagg, J. (1997). Deliberate self-harm in alcohol and drug misusers: patient characteristics and patterns of clinical care. *Drug and Alcohol Review*, **16**, 123–9.
40. Choquet, M. and Ledoux, S. (1994). *Adolescents*: *enquête nationale*. Inserm, Villejuif Cedex.
41. Hawton, K., Rodham, K., Evans, E., *et al.* (2002). Deliberate self harm in adolescents: self-report survey in schools in England. *British Medical Journal*, **325**, 1207–11.
42. Paykel, E.S., Myers, J.K., Lindentall, J.J., *et al.* (1974). Suicidal feelings in the general population: a prevalence study. *The British Journal of Psychiatry*, **124**, 460–9.
43. Kienhorst, C.W.M., de Wilde, E.J., Diekstra, R.F.W., *et al.* (1991). Construction of an index for predicting suicide attempts in depressed adolescents. *The British Journal of Psychiatry*, **159**, 676–82.
44. Rubinstein, J.L., Heeren, T., Housman, D., *et al.* (1989). Suicidal behavior in 'normal' adolescents: risk and protective factors. *The American Journal of Orthopsychiatry*, **59**, 59–71.
45. Bertelote, J.M, Fleischmann, A., De Leo, D., *et al.* (2005). Suicide attempts, plans, and ideation in culturally diverse sites: the WHO SUPRE-MISS community survey. *Psychological Medicine*, **35**, 1–9.
46. Bernal, M., Haro, J.M., Bernert, S., *et al.* (in press). Risk factors for suicidality in Europe: results from the ESEMED study. *Journal of Affective Disorders*.
47. Arensman, E. (1997). Attempted suicide: epidemiology and classification. Unpublished PhD. Dissertation, University of Leiden, The Netherlands.
48. Williams, J.M.G. (1997). *The cry of pain*. Harmondsworth, Penguin.
49. Sinclair, J.M., Crane, C., Hawton, *et al.* (in press). The role of autobiographical memory specificity in deliberate self-harm: correlates and consequences. *Journal of Affective Disorders*.
50. MacLeod, A.K., Rose, G.S., and Williams, J.M.G. (1993). Components of hopelessness about the future in parasuicide. *Cognitive Therapy and Research*, **17**, 441–55.
51. McAuliffe, C., Corcoran, P., Keeley, H.S., *et al.* (2006). Problem-solving ability and repetition of deliberate self-harm: a multicentre study. *Psychological Medicine*, **36**, 45–55.
52. Maris, R.W. (1992). The relationship of nonfatal suicide attempts to completed suicide. In *Assessment and prediction of suicide* (eds. R.W. Maris, A.L. Berman, J.T. Maltsberger, and R.I. Yufit), pp. 362–80. Guilford Press, New York.
53. Kreitman, N. and Casey, P. (1988). Repetition of parasuicide: an epidemiological and clinical study. *The British Journal of Psychiatry*, **153**, 792–800.
54. Hawton, K. and Fagg, J. (1995). Repetition of attempted suicide: the performance of the Edinburgh predictive scales in patients in Oxford. *Archives of Suicide Research*, **1**, 261–72.
55. Hawton, K. and Catalan, J. (1981). Psychiatric management of attempted suicide patients. *British Journal of Hospital Medicine*, **26**, 365–8.
56. Hawton, K. and Fagg, J. (1988). Suicide and other causes of death, following attempted suicide. *The British Journal of Psychiatry*, **152**, 359–66.
57. Nordstrom, P., Samuelsson, M., and Asberg, M. (1995). Survival analysis of suicide risk after attempted suicide. *Acta Psychiatrica Scandinavica*, **91**, 336–40.
58. Suokas, J. and Lonnqvist, J. (1991). Outcome of attempted suicide and psychiatric consultation: risk factors and suicide mortality during a five year follow-up. *Acta Psychiatrica Scandinavica*, **84**, 545–9.
59. Cullberg, J., Wasserman, D., and Stefansson, C.G. (1988). Who commits suicide after a suicide attempt? An 8 to 10 year follow-up in a suburban catchment area. *Acta Psychiatrica Scandinavica*, **77**, 598–603.
60. Hawton, K., Arensman, E., Wasserman, D., *et al.* (1998). Relation between attempted suicide and suicide rates among young people in Europe. *Journal of Epidemiology and Community Health*, **52**, 191–4.
61. Goldacre, M. and Hawton, K. (1985). Repetition of self-poisoning and subsequent death in adolescents who take overdoses. *The British Journal of Psychiatry*, **146**, 395–8.
62. Kerkhof, A.J.F.M. and Arensman, E. (2004). Repetition of attempted suicide: frequent, but hard to predict. In *Suicidal behaviour: theories and findings* (eds. D. De Leo, U. Bille-Brahe, A.J.F.M. Kerkhof, and A. Schmidtke). Hogrefe and Huber, Göttingen.
63. Arensman, E., Kerkhof, A., Dirkzwager, A., *et al.* (1999). Prevalence and risk factors for repeated suicidal behaviour: results from the WHO/EURO multicentre study on parasuicide, 1989–1992 Report University of Leiden, The Netherlands.
64. Van Egmond, M. and Diekstra, R.F.W. (1989). The predictability of suicidal behaviour: the results of a meta-analysis of published studies. In *Suicide and its prevention* (eds. R.F.W. Diekstra, R. Maris, S. Platt, A. Schmidtke, and G. Sonneck), pp. 37–61. Brill, Leiden.
65. Kreitman, N. and Foster, J. (1991). Construction and selection of predictive scales, with special reference to parasuicide. *The British Journal of Psychiatry*, **159**, 185–92.

4.15.3 Biological aspects of suicidal behaviour

J. John Mann and Dianne Currier

Modelling suicidal behaviours

To understand the biological underpinnings of multi-determined behaviours such as suicide and attempted suicide it is necessary to situate them within an explanatory model that can elaborate the causal pathways and interrelations between biological, clinical, genetic, and environmental factors that all play a role in suicidal behaviour. Where possible, such a model should be clinically explanatory, incorporate biological correlates, be testable in both clinical and biological studies, and have some utility in identifying high-risk individuals.

We have proposed a stress–diathesis model of suicidal behaviour wherein exposure to a stressor precipitates a suicidal act in those with the diathesis, or propensity, for suicidal behaviour.[1] Stressors are generally state-dependent factors such as an episode of major depression or adverse life event. The diathesis, we have hypothesized, comprises trait characteristics such as impulsive aggression, and pessimism.[1] Uncovering the biological mechanisms relevant to the stress and the diathesis dimensions of suicidal behaviour will facilitate the identification of both enduring and proximal markers of risk, as well as potential targets for treatment.

One biological correlate of the diathesis for suicidal behaviour appears to be low serotonergic activity. Abnormal serotonergic function may be the result of numerous factors including genetics, early life experience, chronic medical illness, alcoholism or substance use disorder, many of which have been correlated with increased risk for suicidal behaviour. Moreover, serotonergic dysfunction may underlie recurrent mood disorders or behavioural traits that characterize the diathesis, such as aggression and impulsivity. In terms of stress response, the noradrenergic and HPA axis have been the focus of biological studies in suicidal behaviour. This chapter gives an overview of the major neurobiological findings in suicide and attempted suicide, as well as emerging findings from studies of genes related to those systems.

Serotonergic system

Serotonin is involved in brain development, behavioural regulation, modulation of sleep, mood, anxiety, cognition, and memory and is shown to be disturbed in various psychiatric disorders. Serotonergic function is under genetic control and, moreover, deficits in functioning have been shown to be enduring, marking it as a biological trait. The serotonergic system became a target for investigation in relation to suicide when, more than 30 years ago, Asberg and colleagues observed that depressed individuals who had either attempted suicide by violent means or subsequently died by suicide in the study follow-up period were more likely to have lower CSF 5-HIAA levels.[2] Since that time the function of the serotonergic system in suicide and attempted has been examined in many paradigms, and while not all studies agree, there is substantial consensus that individuals who die by suicide, or make serious non-fatal suicide attempts, exhibit a deficiency in CNS serotonin neurotransmission.

Evidence of hypofunction comes from *cerebrospinal fluid* and postmortem studies. 5-hydroxyindoleacetic acid (5-HIAA) is the major metabolite of serotonin and level of CSF 5-HIAA is a guide to serotonin activity in parts of the brain including the prefrontal cortex. There have been over 20 studies of CSF 5-HIAA and suicidal behaviour in mood disorders, and a meta-analysis of prospective studies of 5-HIAA found that in mood disorders lower CSF 5-HIAA increased the chance of death by suicide over fourfold over follow-up periods of 1–14 years.[3]

Multiple *postmortem studies* of suicide, report lower brainstem levels of 5-HIAA and serotonin (5-hydroxytryptamine, 5-HT) (see Mann *et al.* for a review[4]). These deficits in 5-HT or 5-HIAA are observable across diagnostic groups[5] and, despite early reports to the contrary, independent of suicide method. This abnormality appears to be largely specific to the brainstem, and multiple studies have reported no differences between suicides and controls in 5-HT level in other brain regions including the hippocampus, the occipital cortex, the frontal cortex, the temporal cortex, the caudate, the striatum, or the hypothalamus.[4] Serotonin neurone cell bodies are in the brainstem raphe nuclei, while their axons innervate most of the brain including the ventral prefrontal cortex. Morphological analysis of stained serotonin neurones in the brainstem of depressed suicides and non-suicides observed greater cell density in the dorsal raphe nucleus in the suicides[6] suggesting that reduction in serotonin activity is associated with dysfunctional neurones and not with fewer neurones.

Neuroendocrine challenge studies using fenfluramine provide further evidence of anomalous serotonergic function associated with suicidal behaviour. Fenfluramine is a serotonin-releasing drug and a reuptake inhibitor that may also directly stimulate postsynaptic 5-HT receptors. The release of serotonin by fenfluramine causes a measurable increase in serum prolactin levels that is an indirect index of central serotonergic responsiveness. In depressed patients, those with a history of suicide attempts have a more blunted prolactin response to fenfluramine challenge than non-attempters with some evidence that the effect is more strongly related to seriousness of past attempt.[7]

Studies of receptors suggest lower serotonergic transmission in the central nervous system may be accompanied by a compensatory upregulation of some serotonergic postsynaptic receptors such as the 5-HT_{1A} and 5-HT_{2A}, and a decrease in the number of serotonin reuptake sites.[4] There is a reported increase in the concentration of the postsynaptic 5-HT_{2A} receptors in the prefrontal cortex of suicides compared with non-suicides.[8] This increased binding is reflected in more protein and may be due to elevated gene expression in youth suicide.[9] Elevated 5-HT_{2A} binding has also been reported in the amygdala in depressed suicides. In depressed and non-depressed suicides there is evidence that 5-HT_{2A} receptors are upregulated in the dorsal prefrontal cortex but not the rostral prefrontal cortex.[8]

Platelet studies examine 5-HT_{2A} in living subjects with respect to non-fatal suicide attempt. 5-HT_{2A} receptors, serotonin reuptake sites, and serotonin second messenger systems are present in blood platelets, and changes in these platelet measurements may reflect similar changes in the CNS. Multiple studies have reported higher platelet 5-HT_{2A} receptor numbers in suicide attempters compared with non-attempters and healthy controls.[11]

Studies of second messengers indicate impaired 5-HT_{2A} receptor mediated signal transduction in the prefrontal cortex of suicides, [12] and in platelets 5-HT_{2A} receptor responsivity is significantly blunted in patients with major depression who have made high-lethality suicide attempts compared to depressed patients who have made low-lethality suicide attempts.[13] The implications of such a defect in signal transduction, if present in the brain, would be that although there may be greater density of 5-HT_{2A} receptors, the signal transduced by 5-HT_{2A} receptor activation may be blunted, which would compound deficient serotonergic input as seen in the lower levels of brainstem serotonin and/or 5-HIAA in suicide victims.

Some *postmortem studies* of the postsynaptic 5-HT_{1A} receptor report higher binding in prefrontal cortex and more rostral segments of raphe nuclei, and lower binding in more caudal raphe nuclei, hippocampus, prefrontal cortex, and temporal cortex.[7] Less 5-HT_{1A} autoreceptor gene expression is also reported in the dorsal raphe[14] and would favour higher serotonin neurone firing rates.

Postmortem studies of depressed suicides report fewer 5-HT transporters in prefrontal cortex, hypothalamus, occipital cortex, and brainstem.[15] Moreover, in suicides this deficit appears localized to the ventromedial prefrontal cortex, whereas depressed individuals who died of other causes had lower binding throughout the prefrontal cortex.[16]

The emerging picture from postmortem studies of greater 5-HT_{2A} receptor binding in the frontal cortex of depressed individuals who die by suicide, fewer brainstem 5-HT_{1A} autoreceptors, and fewer serotonin transporters in the cortex, as well as findings of greater tryptophan hydroxylase (the rate-limiting step in serotonin synthesis) immunoreactivity in serotonin nuclei in the brainstem[17] all point to homeostatic changes designed to increase deficient serotonergic transmission evidenced by low 5-HIAA in CSF and

brain, low 5-HT and 5-HIAA in brainstem, and blunted prolactin response to fenfluramine challenge.

Serotonergic dysfunction and suicide endophenotypes. Increased aggression has been associated with suicide and more highly lethal suicide attempts and impulsivity has shown a stronger relationship to non-fatal suicide attempts.[18] Impulsive aggressive traits are potentially part of the diathesis for suicidal behaviour.[1] Reduced activity of the serotonin system has been implicated in impulsive violence and aggression in studies in a variety of paradigms including: Low CSF 5-HIAA in individuals with a lifetime history of aggressive behaviour with personality and other psychiatric disorders;[19, 20] a blunted prolactin response to serotonin-releasing agent fenfluramine in personality disorder patients,[21, 22] and; greater platelet 5-HT_{2A} binding correlated with aggressive behaviour in personality and other psychiatric disorder patients.[23, 24] In a postmortem study of aggression, suicidal behaviour, and serotonergic function a positive relationship between lifetime history of aggression scores and 5-HT_{2A} binding in several regions of prefrontal cortex of individuals who had died by suicide was found.[25]

Positron emission tomography (PET) studies have shown a deficient response to serotonergic challenge in the orbitofrontal cortex, medial frontal, and cingulate regions in individuals with impulsive aggression compared to controls[26, 27] and lower serotonin transporter binding in the anterior cingulate cortex in impulsive aggressive individuals compared to healthy controls.[28] The prefrontal cortex is important in the inhibitory control of behaviour, including impulsive and aggressive behaviour.[29] Thus aggressive/impulsive traits, related to serotonergic dysfunction, are potentially an aspect of the diathesis for suicidal behaviour, whereby aggressive/suicidal behaviours is manifested in response to stressful circumstances or powerful emotions. This tendency might be conceived of as a diminution in natural inhibitory circuits, or as a volatile cognitive decision style.

Noradrenergic system

Within the stress–diathesis model of suicidal behaviour, it is the confluence of stressful events with the diathesis that is thought to precipitate a suicidal act. Thus, investigating the functioning of stress response systems in suicidal individuals is important for elucidating neurobiological concomitants of suicidal behaviour and identifying targets for preventative intervention. The noradrenergic system and the HPA axis are two key stress response systems.

The majority of norepinephrine neurones in the brain are located in the brainstem locus coeruleus. Postmortem studies of suicides have documented fewer noradrenergic neurones in the locus coeruleus.[30] There are also indications of cortical noradrenergic overactivity including lower alpha and high-affinity $beta_1$-adrenergic receptor binding,[31] and lower β-adrenoceptor density and $alpha_2$–adrenergic binding in the prefrontal cortex in individuals who died by suicide.[23] There is some, but not unanimous, evidence from prospective studies of lower levels of CSF 3-methoxy-4-hydroxphenylglycol (MHPG), a metabolite of noradrenaline, in future suicides,[33] although not in those making non-fatal suicide attempts.[34]

Fewer noradrenergic neurones observed in depressed suicides may indicate a lower functional reserve of the noradrenergic system, which if accompanied by an exaggerated stress response with greater release of noradrenaline may result in norepinephrine depletion leading to depression and hopelessness, both of which are contributory factors to suicidal behaviour.

Noradrenergic and HPA axis responses to stress in adulthood appears to be greater in those reporting an abusive experience in childhood.[35] Such individuals are potentially at greater risk in adulthood for major depression and suicidal behaviour. Childhood abuse may be associated with increased risk for depression and suicidal behaviour because of a dysfunctional stress response both via the noradrenergic system and the HPA axis, and secondary effects of norepinephrine depletion and elevated cortisol levels. There is interaction between the noradrenergic system and the stress response activity of the HPA axis with reciprocal neural connections between corticotropin-releasing hormone neurones in the hypothalamic paraventricular nucleus and noradrenergic neurones in human brainstem and the locus coeruleus.[36]

The hypothalamic-pituitary-adrenal (HPA) axis

The hypothalamic-pituitary-adrenal axis is a major stress response system. Major depression is associated with hyperactivity of the HPA axis,[37] and suicidal patients in diagnostically heterogeneous populations exhibit HPA axis abnormalities, most commonly failure to suppress cortisol normally after dexamethasone.[33] We found most future suicides were dexamethasone suppression test (DST) non-suppressors.[33] In mood disorders, DST non-suppressors had a 4.5-fold greater risk of dying by suicide compared with suppressors.[3] Moreover, non-suppression may be characteristic of more serious attempts that result in greater medical damage[38, 39] or the use of violent method in the suicide attempt.[40] In other indices of HPA axis function suicide attempters had attenuated plasma cortisol responses to fenfluramine although that may indicate less serotonin release and not an HPA abnormality,[41, 42] and lower CSF corticotropin-releasing hormone (CRH) compared to non-attempters,[43] though not all studies agree.

Larger pituitary and larger adrenal gland volumes are reported in depressed suicides,[44, 45] and fewer CRH-binding sites in the prefrontal cortex of depressed suicide victims which may mean receptor downregulation due to elevated CRF release.[46]

As with the noradrenergic system, early life adversity appears to have lasting effects on stress response in the HPA axis in adulthood. Abnormalities in HPA axis function have been implicated in poor response to antidepressant treatment, and greater likelihood of relapse in major depression, both of which increase the risk for suicidal acts.[33] Increased anxiety and agitation are another potential pathway whereby abnormal stress response, in both the noradrenergic and HPA axis, contributes to risk for suicidal behaviour.

Other biologic systems

Abnormality in the dopaminergic system has been reported in depressive disorders,[47] however studies of dopaminergic function and suicidal behaviour are relatively few and inconclusive.[48] Low dihydroxyphenylacetic acid levels, indicative of reduced dopamine turnover, in the caudate, putamen, and nucleus accumbens are reported in depressed suicides,[49] although the same group of investigators found no difference in number or affinity of the dopamine

transporters.[50] Accordingly, it is unlikely that the reduced dopamine turnover initially observed in depressed suicides is a result of decreased dopaminergic innervation of those regions. Prospective studies disagree as to whether CSF HVA predicts suicidal behaviour.[51–53]

There is a well-documented relationship between thyroid dysfunction and depression[54] and some studies link thyroid function and suicide. Abnormal thyroid-stimulating hormone (TSH) response to thyrotropin-releasing hormone (TRH) has been observed in individuals who died by suicide in a follow-up study. [55] Abnormal TSH response to challenge tests has also been associated with poor response to antidepressant treatment and a higher relapse rate, which may increase risk for suicidal behaviour.[56]

Neurotrophins are involved in brain development and growth, neuronal functioning, and synaptic plasticity. Lower protein levels and gene expression of brain-derived neurotrophic factor (BDNF) in the prefrontal cortex and hippocampus,[57,58] and less mRNA of nerve growth factor, neurotrophin 3 and neurotrophin 4/5 in the hippocampus[59] are reported postmortem in suicides. Lower plasma BDNF has been reported in MDD suicide attempters compared to MDD non-attempters and healthy controls.[60]

Suicide is more common in groups with very low cholesterol levels or after cholesterol lowering by diet (see[61] for a review). This relationship between cholesterol and suicide may be mediated by serotonergic function, as studies of non-human primates on a low-fat diet found lower serotonergic activity and increased aggressive behaviours.[62] Long chain polyunsaturated fatty acids, particularly omega-3, may also be a mediating factor in the relationship between low cholesterol and increased risk for depression and suicide.[63] Lower docosahexaenoic acid percentage of total plasma polyunsaturated fatty acids and a higher omega-6/omega-3 ratio predicted depressed individuals who made a suicide attempt during a 2-year follow-up,[64] and lower eicosapentaenoic acid is found in red blood cells of suicide attempters compared to controls.[65]

Neurobiology, genetics, and suicidal behaviour

Family, twin, and adoption studies support a genetic contribution to suicidal behaviour independent of psychiatric disorder (see Brent and Mann for a review),[66] and genetic studies have sought to determine the responsible genes for suicide and suicide attempt though linkage and SNP association studies. Candidate genes for most studies were selected based on evidence from neurobiological studies in suicide, as a result of which the serotonergic system has been most extensively investigated. A tri-allelic polymorphism in the serotonin transporter promotor has two alleles with lower transcriptional activity and fewer transporters. In varied psychiatric populations, despite some negative findings, the S, or more common lower-expressing allele, has been associated with suicide and with suicide attempts, particularly violent or high-lethality attempts.[67] Functional MRI studies find greater amygdala activation in individuals with the SS genotype when they are exposed to negative stimuli such as angry or fearful faces, negative words, or aversive pictures (see Brown and Hariri 2006 for a review).[68] The amygdala is densely innervated by serotonergic neurones and 5-HT receptors are abundant, and plays a central role in emotional regulation and memory. Excessive responses to emotionally negative events such as abuse, may be over-encoded and contribute to stress-sensitivity in adulthood and thereby to major depression after stress and even suicidal behaviour.

Other genetic studies of the serotonergic system including the 5-HT_{1A}, 5-HT_{2A}, 5-HT_{1B} and other serotonin receptors have largely reported negative results, although there have been some positive findings for the 5-HT_{2A} 102C allele and attempted suicide or suicidal ideation.[67] For tryptophan hydroxylase (TPH1 and TPH2 are two forms of TPH with TPH1 only expressed in the brain during development), associations are reported with suicide and suicide attempt and TPH1 SNPs, however multiple negative findings have also been reported,[67] while haplotype and SNP studies suggest the involvement of the TPH2 gene in suicide and suicide attempt, however again not all studies agree.[67] Monoamine oxidase (MAO-A) plays a key role in metabolism of amines. Low MAO activity results in elevated levels of serotonin, norepinephrine, and dopamine in the brain. The MAO-A gene has functional variable number tandem repeat however no association has been found between this uVNTR and suicidal behaviour, although there is some indication that it may be related to aggression[67] and it is linked to the impact of adversity in childhood on adult antisocial behaviour and trait impulsiveness.[69,70]

Genetic studies in dopaminergic system, noradrenergic system, BDNF, and GABA are few and generally negative,[67] although there are reports of positive association of the catechol-*O*-methyltransferase (COMT), a major catecholamine-catabolic enzyme, gene in Finnish and Caucasian suicide attempters[71] and in Japanese suicides.[72] Inconsistent findings in genetic studies of suicidal behaviour may be due to the complexity of the suicide phenotype, gene–gene interactions, the presence of multiple psychiatric disorders, population racial differences, possible epigenetic effects, and the influence of gene/environment interactions. Nonetheless, new microarray technologies that test expression of thousands of genes simultaneously allowing better gene coverage, and haplotype mapping approaches offer promise for future investigation. Other options include examining more basic endophenotypes such as mood regulation and decision-making.

Genes and environment

Early life stress in conjunction with genetic vulnerability can have enduring effects into adulthood and affect psychopathology and the functioning of biological systems including the serotonergic and stress-response systems (see Mann and Currier 2006).[73] For example monkeys exposed to maternal deprivation in infancy and having the 5-HTTLPR lower expressing S allele in the serotonin transporter gene manifest a lowering of CSF 5-HIAA that persists into adulthood.[74] In 6-month-old macaque monkeys exposed to social stress, those with the S allele had a higher ACTH response, an HPA axis hormone related to stress response, compared with those without that allele and to S allele animals who were maternally reared.[75] Thus the low-expressing S allele not only increased vulnerability to stress in development, but early life stress may further interact with genotype to lower serotonergic function and to increase sensitivity to stressful events later in life, both of which are risk factors for suicidal behaviour.

In human studies, individuals who had experienced childhood maltreatment, those with the low-expressing S allele were at risk for

suicidal ideation and suicide attempt,[76] and those with a lower expressing variant of the MAO-A gene were more likely to manifest antisocial behaviour and more impulsivity as adults.[69, 70]

Future directions

There is much still to be learned about the biologic aetiology of suicidal behaviour and the pathways and mechanisms through which biologic dysfunction is involved in suicidal acts. New techniques for imaging the brain, identification of basic intermediate phenotypes and denser gene markers will contribute to elucidating the biological factors and mechanisms involved in suicide and attempted suicide, and identifying potential targets for prevention.

Further information

Mann, J.J. (2003). Neurobiology of suicidal behaviour. *Nature Reviews Neuroscience*, **4**(10), 819–28.

Pandey, G.N. (1997). Altered serotonin function in suicide. Evidence from platelet and neuroendocrine studies. *Annals of the New York Academy of Sciences*, **836**, 182–200.

Mann, J.J. and Currier, D. (2007). A review of prospective studies of biologic predictors of suicidal behavior in mood disorders. *Archives of Suicide Research: Official Journal of the International Academy for Suicide Research*, **11**(1), 3–16.

Bondy, B., Buettner, A., and Zill, P. (2006). Genetics of suicide. *Molecular Psychiatry*, **11**(4), 336–51.

References

1. Mann, J.J., Waternaux, C., Haas, G.L., *et al.* (1999). Toward a clinical model of suicidal behavior in psychiatric patients. *The American Journal of Psychiatry*, **156**(2), 181–9.
2. Åsberg, M., Thorén, P., Träskman, L., Bertilsson, L., and Ringberger, V. (1976). "Serotonin depression"—A biochemical subgroup within the affective disorders? *Science*, **191**, 478–80.
3. Mann, J.J., Currier, D., Stanley, B., *et al.* (2006). Can biological tests assist prediction of suicide in mood disorders? *The International Journal of Neuropsychopharmacology*, **9**(4), 465–74.
4. Mann, J.J., Underwood, M.D., and Arango, V. (1996). Postmortem studies of suicide victims. In *Biology of schizophrenia and affective disease* (ed. S.J. Watson), pp. 197–220. American Psychiatric Press, Washington, DC.
5. Mann, J.J. (1998). The neurobiology of suicide. *Nature Medicine*, **4**(1), 25–30.
6. Arango, V., Underwood, M.D., Gubbi, A.V., *et al.* (1995). Localized alterations in pre- and postsynaptic serotonin binding sites in the ventrolateral prefrontal cortex of suicide victims. *Brain Research*, **688**(1–2), 121–33.
7. Kamali, M., Oquendo, M.A., and Mann, J.J. (2001). Understanding the neurobiology of suicidal behavior. *Depression and Anxiety*, **14**(3), 164–76.
8. Stockmeier, C.A. (2003). Involvement of serotonin in depression: evidence from postmortem and imaging studies of serotonin receptors and the serotonin transporter. *Journal of Psychiatric Research*, **37**(5), 357–73.
9. Pandey, G.N., Dwivedi, Y., Rizavi, H.S., *et al.* (2002). Higher expression of serotonin 5-HT(2A) receptors in the postmortem brains of teenage suicide victims. *The American Journal of Psychiatry*, **159**(3), 419–29.
10. Hrdina, P.D., Demeter, E., Vu, T.B., *et al.* (1993). 5-HT uptake sites and 5-HT_2 receptors in brain of antidepressant-free suicide victims/depressives: increase in 5-HT_2 sites in cortex and amygdala. *Brain Research*, **614**, 37–44.
11. Pandey, G.N. (1997). Altered serotonin function in suicide. Evidence from platelet and neuroendocrine studies. *Annals of the New York Academy of Sciences*, **836**, 182–200.
12. Pandey, G.N., Dwivedi, Y., Pandey, S.C., *et al.* (1999). Low phosphoinositide-specific phospholipase C activity and expression of phospholipase C beta1 protein in the prefrontal cortex of teenage suicide subjects. *The American Journal of Psychiatry*, **156**(12), 1895–901.
13. Malone, K.M., Ellis, S.P., Currier, D., *et al.* (2007). Platelet 5-HT2A receptor subresponsivity and lethality of attempted suicide in depressed in-patients. *The International Journal of Neuropsychopharmacology*, **10**(3): 335–430.
14. Arango, V., Underwood, M.D., Boldrini, M., *et al.* (2001). Serotonin 1A receptors, serotonin transporter binding and serotonin transporter mRNA expression in the brainstem of depressed suicide victims. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **25**(6), 892–903.
15. Purselle, D.C. and Nemeroff, C.B. (2003). Serotonin transporter: a potential substrate in the biology of suicide. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **28**(4), 613–19.
16. Mann, J.J., Huang, Y.Y., Underwood, M.D., *et al.* (2000). A serotonin transporter gene promoter polymorphism (5-HTTLPR) and prefrontal cortical binding in major depression and suicide. *Archives of General Psychiatry*, **57**(8), 729–38.
17. Boldrini, M., Underwood, M.D., Mann, J.J., *et al.* (2005). More tryptophan hydroxylase in the brainstem dorsal raphe nucleus in depressed suicides. *Brain Research*, **1041**(1), 19–28.
18. Oquendo, M.A., Galfalvy, H., Russo, S., *et al.* (2004). Prospective study of clinical predictors of suicidal acts after a major depressive episode in patients with major depressive disorder or bipolar disorder. *The American Journal of Psychiatry*, **161**(8), 1433–41.
19. Brown, G.L. and Goodwin, F.K. (1986). Cerebrospinal fluid correlates of suicide attempts and aggression. *Annals of the New York Academy of Sciences*, **487**, 175–88.
20. Stanley, B., Molcho, A., Stanley, M., *et al.* (2000). Association of aggressive behavior with altered serotonergic function in patients who are not suicidal. *The American Journal of Psychiatry*, **157**(4), 609–14.
21. Coccaro, E.F., Siever, L.J., Klar, H.M., *et al.* (1989). Serotonergic studies in patients with affective and personality disorders. Correlates with suicidal and impulsive aggressive behavior. *Archives of General Psychiatry*, **46**, 587–99.
22. New, A.S., Trestman, R.F., Mitropoulou, V., *et al.* (2004). Low prolactin response to fenfluramine in impulsive aggression. *Journal of Psychiatric Research*, **38**(3), 223–30.
23. Coccaro, E.F., Kavoussi, R.J., Sheline, Y.I., *et al.* (1997). Impulsive aggression in personality disorder correlates with platelet 5-HT_{2A} receptor binding. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **16**(3), 211–16.
24. McBride, P.A., Brown, R.P., DeMeo, M., *et al.* (1994). The relationship of platelet 5-HT_2 receptor indices to major depressive disorder, personality traits, and suicidal behavior. *Biological Psychiatry*, **35**, 295–308.
25. Oquendo, M.A., Russo, S.A., Underwood, M.D., *et al.* (2006). Higher postmortem prefrontal 5-HT2A receptor binding correlates with lifetime aggression in suicide. *Biological Psychiatry*, **59**(3), 235–43.
26. Siever, L.J., Buchsbaum, M.S., New, A.S., *et al.* (1999). D,1-Fenfluramine response in impulsive personality disorder assessed with [^{18}F]fluorodeoxyglucose positron emission tomography. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **20**(5), 413–23.
27. New, A.S., Hazlett, E.A., Buchsbaum, M.S., *et al.* (2002). Blunted prefrontal cortical 18fluorodeoxyglucose positron emission tomography response to meta-chlorophenylpiperazine in impulsive aggression. *Archives of General Psychiatry*, **59**(7), 621–9.

28. Frankle, W.G., Lombardo, I., New, A.S., *et al.* (2005). Brain serotonin transporter distribution in subjects with impulsive aggressivity: a positron emission study with [^{11}C]McN 5652. *The American Journal of Psychiatry*, **162**(5), 915–23.
29. de Almeida, R.M., Rosa, M.M., Santos, D.M., *et al.* (2006). 5-HT(1B) receptors, ventral orbitofrontal cortex, and aggressive behavior in mice. *Psychopharmacology*, **185**(4), 441–50.
30. Arango, V., Underwood, M.D., and Mann, J.J. (1996). Fewer pigmented locus coeruleus neurons in suicide victims: preliminary results. *Biological Psychiatry*, **39**, 112–20.
31. Arango, V., Ernsberger, P., Sved, A.F., *et al.* (1993). Quantitative autoradiography of a_1- and a_2-adrenergic receptors in the cerebral cortex of controls and suicide victims. *Brain Research*, **630**, 271–82.
32. De Paermentier, F., Cheetham, S.C., Crompton, M.R., *et al.* (1990). Brain b-adrenoceptor binding sites in antidepressant-free depressed suicide victims. *Brain Research*, **525**, 71–7.
33. Mann, J.J. and Currier, D. (2007). A review of prospective studies of biologic predictors of suicidal behavior in mood disorders. *Archives of Suicide Research: Official Journal of the International Academy for Suicide Research*, **11**(1), 3–16.
34. Lester, D. (1995). The concentration of neurotransmitter metabolites in the cerebrospinal fluid of suicidal individuals: a meta-analysis. *Pharmacopsychiatry*, **28**(2), 45–50.
35. Heim, C. and Nemeroff, C.B. (2001). The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biological Psychiatry*, **49**(12), 1023–39.
36. Austin, M.C., Rice, P.M., Mann, J.J., *et al.* (1995). Localization of corticotropin-releasing hormone in the human locus coeruleus and pedunculopontine tegmental nucleus: an immunocytochemical and in situ hybridization study. *Neuroscience*, **64**(3), 713–27.
37. Carroll, B.J., Feinberg, M., Greden, J.F., *et al.* (1981). A specific laboratory test for the diagnosis of melancholia. Standardization, validation, and clinical utility. *Archives of General Psychiatry*, **38**, 15–22.
38. Norman, W.H., Brown, W.A., Miller, I.W., *et al.* (1990). The dexamethasone suppression test and completed suicide. *Acta Psychiatrica Scandinavica*, **81**(2), 120–5.
39. Coryell, W. (1990). DST abnormality as a predictor of course in major depression. *Journal of Affective Disorders*, **19**(3), 163–9.
40. Roy, A. (1992). Hypothalamic-pituitary-adrenal axis function and suicidal behavior in depression. *Biological Psychiatry*, **32**, 812–16.
41. Duval, F., Mokrani, M.C., Correa, H., *et al.* (2001). Lack of effect of HPA axis hyperactivity on hormonal responses to d- fenfluramine in major depressed patients: implications for pathogenesis of suicidal behaviour. *Psychoneuroendocrinology*, **26**(5), 521–37.
42. Malone, K.M., Corbitt, E.M., Li, S., *et al.* (1996). Prolactin response to fenfluramine and suicide attempt lethality in major depression. *The British Journal of Psychiatry: The Journal of Mental Science*, **168**, 324–9.
43. Brunner, J., Stalla, G.K., Stalla, J., *et al.* (2001). Decreased corticotropin-releasing hormone (CRH) concentrations in the cerebrospinal fluid of eucortisolemic suicide attempters. *Journal of Psychiatric Research*, **35**(1), 1–9.
44. Szigethy, E., Conwell, Y., Forbes, N.T., *et al.* (1994). Adrenal weight and morphology in victims of completed suicide. *Biological Psychiatry*, **36**(6), 374–80.
45. Dumser, T., Barocka, A., and Schubert, E. (1998). Weight of adrenal glands may be increased in persons who commit suicide. *The American Journal of Forensic Medicine and Pathology: Official Publication of the National Association of Medical Examiners*, **19**(1), 72–6.
46. Nemeroff, C.B., Owens, M.J., Bissette, G., *et al.* (1988). Reduced corticotropin releasing factor binding sites in the frontal cortex of suicide victims. *Archives of General Psychiatry*, **45**, 577–9.
47. Dailly, E., Chenu, F., Renard, C.E., *et al.* (2004). Dopamine, depression and antidepressants. *Fundamental and Clinical Pharmacology*, **18**(6), 601–7.
48. Mann, J.J. (2003). Neurobiology of suicidal behaviour. *Nature Reviews Neuroscience*, **4**(10), 819–28.
49. Bowden, C., Cheetham, S.C., Lowther, S., *et al.* (1997). Reduced dopamine turnover in the basal ganglia of depressed suicides. *Brain Research*, **769**(1), 135–40.
50. Bowden, C., Theodorou, A.E., Cheetham, S.C., *et al.* (1997). Dopamine D_1 and D_2 receptor binding sites in brain samples from depressed suicides and controls. *Brain Research*, **752**, 227–33.
51. Roy, A., De Jong, J., and Linnoila, M. (1989). Cerebrospinal fluid monoamine metabolites and suicidal behavior in depressed patients. A 5-year follow-up study. *Archives of General Psychiatry*, **46**, 609–12.
52. Engstrom, G., Alling, C., Blennow, K., *et al.* (1999). Reduced cerebrospinal HVA concentrations and HVA/5-HIAA ratios in suicide attempters. Monoamine metabolites in 120 suicide attempters and 47 controls. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, **9**(5), 399–405.
53. Placidi, G.P., Oquendo, M.A., Malone, K.M., *et al.* (2001). Aggressivity, suicide attempts, and depression: relationship to cerebrospinal fluid monoamine metabolite levels. *Biological Psychiatry*, **50**(10), 783–91.
54. Jackson, I.M. (1998). The thyroid axis and depression. *Thyroid: Official Journal of the American Thyroid Association*, **8**(10), 951–6.
55. Linkowski, P., Van Wettere. J.P., Kerkhofs, M., *et al.* (1984). Violent suicidal behavior and the thyrotropin-releasing hormone-thyroid-stimulating hormone test: a clinical outcome study. *Neuropsychobiology*, **12**(1), 19–22.
56. Targum, S.D. (1984). Persistent neuroendocrine dysregulation in major depressive disorder: a marker for early relapse. *Biological Psychiatry*, **19**(3), 305–18.
57. Dwivedi, Y., Rizavi, H.S., Conley, R.R., *et al.* (2003). Altered gene expression of brain-derived neurotrophic factor and receptor tyrosine kinase B in postmortem brain of suicide subjects. *Archives of General Psychiatry*, **60**(8), 804–15.
58. Karege, F., Vaudan, G., Schwald, M., *et al.* (2005). Neurotrophin levels in postmortem brains of suicide victims and the effects of antemortem diagnosis and psychotropic drugs. *Brain Research Molecular Brain Research*, **136**(1–2), 29–37.
59. Dwivedi, Y., Mondal, A.C., Rizavi, H.S., *et al.* (2005). Suicide brain is associated with decreased expression of neurotrophins. *Biological Psychiatry*, **58**(4), 315–24.
60. Kim, Y.K., Lee, H.P., Won, S.D., *et al.* (2007). Low plasma BDNF is associated with suicidal behavior in major depression. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **31**(1), 78–85.
61. Golomb, B.A. (1998). Cholesterol and violence: is there a connection? *Annals of Internal Medicine*, **128**, 478–87.
62. Muldoon, M.F., Rossouw, J.E., Manuck, S.B., *et al.* (1993). Low or lowered cholesterol and risk of death from suicide and trauma. *Metabolism: Clinical and Experimental*, **42**(Suppl. 1), 45–56.
63. Brunner, J., Parhofer, K.G., Schwandt, P., *et al.* (2002). Cholesterol, essential fatty acids, and suicide. *Pharmacopsychiatry*, **35**(1), 1–5.
64. Sublette, M.E., Hibbeln, J.R., Galfalvy, H., *et al.* (2006). Omega-3 polyunsaturated essential fatty acid status as a predictor of future suicide risk. *American Journal of Psychiatry*, **163**(6), 1100–2.
65. Huan, M., Hamazaki, K., Sun, Y., *et al.* (2004). Suicide attempt and *n*–3 fatty acid levels in red blood cells: a case control study in China. *Biological Psychiatry*, **56**(7), 490–6.
66. Brent, D.A. and Mann, J.J. (2005). Family genetic studies, suicide, and suicidal behavior. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, **133**(1), 13–24.
67. Bondy, B., Buettner, A., and Zill, P. (2006). Genetics of suicide. *Molecular Psychiatry*, **11**(4), 336–51.

68. Brown, S.M. and Hariri, A.R. (2006). Neuroimaging studies of serotonin gene polymorphisms: exploring the interplay of genes, brain, and behavior. *Cognitive Affective and Behavioral Neuroscience*, **6**(1), 44–52.
69. Caspi, A., McClay, J., Moffitt, T.E., *et al.* (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, **297**(5582), 851–4.
70. Huang, Y.Y., Cate, S.P., Battistuzzi, C., *et al.* (2004). An association between a functional polymorphism in the monoamine oxidase a gene promoter, impulsive traits and early abuse experiences. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **29**(8), 1498–505.
71. Nolan, K.A., Volavka, J., Czobor, P., *et al.* (2000). Suicidal behavior in patients with schizophrenia is related to COMT polymorphism. *Psychiatric Genetics*, **10**(3), 117–24.
72. Ono, H., Shirakawa, O., Nushida, H., *et al.* (2004). Association between catechol-*O*-methyltransferase functional polymorphism and male suicide completers. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **29**(7), 1374–7.
73. Mann, J.J. and Currier, D. (2006). Effects of genes and stress on the neurobiology of depression. *International Review of Neurobiology*, **73**, 153–89.
74. Bennett, A.J., Lesch, K.P., Heils, A., *et al.* (2002). Early experience and serotonin transporter gene variation interact to influence primate CNS function. *Molecular Psychiatry*, **7**(1), 118–22.
75. Barr, C.S., Newman, T.K., Shannon, C., *et al.* (2004). Rearing condition and rh5-HTTLPR interact to influence limbic-hypothalamic-pituitary-adrenal axis response to stress in infant macaques. *Biological Psychiatry*, **55**(7), 733–8.
76. Caspi, A., Sugden, K., Moffitt, T.E., *et al.* (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, **301**(5631), 386–9.

4.15.4 Treatment of suicide attempters and prevention of suicide and attempted suicide

Keith Hawton and Tatiana Taylor

Introduction

In considering the treatment and prevention of suicidal behaviour account should be taken of recent trends in suicide and attempted suicide, particularly in individual countries. These have been reviewed in other chapters in this section. The term 'attempted suicide' is used in this chapter and includes any act of non-fatal self-poisoning or self-injury, irrespective of motive or intention.

Suicide prevention has been incorporated within the World Health Organization Health for All strategy[1] and has received substantial support from the United Nations.[2] Furthermore, in recent years several countries have developed national suicide prevention programmes. Increased suicide rates in young people have probably acted as a stimulus behind this trend. However, suicide rates in most countries remain higher in older populations and prevention programmes must include this increasingly larger sector of society.

Treatment of suicide attempters

Suicide attempts occur for a wide range of reasons. In many cases the primary aim is not death but some other outcome, such as demonstrating distress to other people, seeking a change in other people's behaviour or temporary escape.[3] This means that a broad range of treatments are required since the needs of individual patients will vary widely (Table 4.15.4.1).

Factors relevant to treatment needs in suicide attempts

(a) Repetition of attempts and risk of suicide

Repetition of attempts is common, with 15 to 25 per cent repeating suicidal acts within a year, and is associated with a greater risk of eventual suicide.[4] The frequency of suicide following attempted suicide varies from country to country,[5] depending on the overall characteristics of the patient population and the rate of suicide in the general population. Prevention of repetition of suicidal behaviour and especially of suicide is a major aim in treating suicide attempters.

(b) Psychiatric and personality disorders

A range of psychiatric disorders are found in suicide attempters.[6] Depression and alcohol abuse are particularly common. In addition, substantial proportions of patients have personality disorders. While treatment directed at the underlying causes of such disorders, where possible, will be important in managing attempted suicide patients, often the disorders themselves will require specific treatment.

(c) Life events and difficulties

Certain problems are particularly common in suicide attempters, including difficulties in interpersonal relationships, especially with partners and with other family members, employment problems, particularly in males, and financial difficulties. Life events, especially disruption in a relationship with a partner, frequently precede suicidal acts.[7]

(d) Poor problem-solving skills

Many suicide attempters have difficulties in problem-solving, particularly in dealing with difficulties in interpersonal relationships.[8] These difficulties are more marked in suicide attempters than in patients with psychiatric disorders who have not carried out a suicidal act.

Table 4.15.4.1 Factors relevant to treatment needs in suicide attempters

Risk of repetition of attempts
Risk of suicide
Psychiatric disorder (especially depression and substance abuse)
Personality disorders
Life events and difficulties
Poor problem-solving skills
Impulsivity and aggression
Hopelessness
Low self-esteem
Motivational problems and poor compliance with treatment

(e) Impulsivity and aggression

There is a strong link between suicidal behaviour and both impulsivity and aggression. There is also accumulating evidence that hypofunction of brain serotonergic systems is linked to aggression (and possibly impulsivity) and also to suicidal behaviour[9] (see Chapter 4.15.3). It is unclear whether this represents a state phenomenon associated with psychiatric disturbance or a trait phenomenon, but current evidence points towards the latter.

(f) Hopelessness and low self-esteem

Hopelessness, or pessimism about the future, which has been shown to be a key factor linking depression with suicidal acts, is an important predictor of repetition of suicidal behaviour, and a risk factor for eventual suicide.[10] Low self-esteem is another important characteristic associated with suicidal behaviour. There is likely to be a link between low self-esteem and a tendency to experience hopelessness when faced by adverse circumstances.

(g) Motivational problems and poor compliance with treatment

Management of suicide attempters is complicated by the fact that some patients appear to be poorly motivated to engage in aftercare. This is also likely to affect compliance with treatments. The style of organization of general hospital psychiatric services (including continuity of care) and the attitudes of clinical staff may be important factors determining whether patients engage in aftercare.

General overview of treatments

Treatments for suicide attempters include both psychosocial and pharmacological approaches. While these are considered separately below, in some patients both will be appropriate. This might be the case, for example, if a patient suffers from depression with biological features in the setting of employment and financial difficulties, when treatment with an antidepressant might be combined with problem-solving therapy.

Psychosocial treatments

A range of psychosocial therapies have been evaluated in suicide attempters in randomized controlled clinical trials. The efficacy of these approaches has been examined in a systematic review of the worldwide literature.[11] The findings from this review and some further studies are summarized below.

(a) Problem-solving

Meta-analysis of the results of trials that have been conducted so far to evaluate the effectiveness of brief problem-solving therapy (see Chapter 6.3.1) compared with treatment as usual indicates a trend towards reduction of repetition of self-harm episodes, but the total numbers of subjects and trials have precluded a definitive result. However, evidence of other positive outcomes, such as reduced levels of depression and hopelessness and improvement in problems, has been convincingly demonstrated in these studies.[12] This approach is useful, either used alone or in the context of other treatment. It is reasonably easily taught and can be used by clinicians from different professional backgrounds.

(b) Psychotherapy

Cognitive behaviour therapy, combined with care management has recently been shown to be effective in reducing frequency of suicide attempts and in producing other positive outcomes.[13] A brief psychological intervention combined with provision of a treatment manual seems to be less effective in the treatment of patients with repeated attempts.[14]

Two trials have been conducted in which an intensive form of psychological treatment known as dialectical behaviour therapy was evaluated.[15,16] Female patients with borderline personality disorders who had a history of repeated self-harm were offered a year of individual and group cognitive behavioural therapy aimed at addressing the patients' problems of motivation and strengthening their behavioural skills, particularly in relation to interpersonal difficulties. Compared with routine care this approach seems to result in a reduction in repetition of self-harm as well as a number of other positive outcomes. Further evaluation of this approach is required to determine if it is effective in male patients and in adolescents, and whether it can be delivered in an abbreviated form. While it is a labour-intensive approach, it appears to be helpful for what is a particularly difficult group of patients.

(c) Outreach

Several trials have been conducted to assess the impact of community outreach, either for all patients or for those that have not attended treatment sessions. Some of these studies have included relatively intensive treatment programmes. Overall these studies indicate that some form of outreach may improve outcome in terms of reducing repetition of attempted suicide. In one study, nurse home visits to encourage non-attending participants to attend outpatient appointments resulted in a significantly greater number of appointments attended as compared to the control group and there was a near significant reduction in the rate of repetition of suicide attempts during the year after study entry.[17] In other studies, telephone contact,[18] and contacting patients regularly by post[19] have also produced promising results. Outreach combined with specific treatment may be useful, perhaps reserved for those who are poorly compliant with aftercare.

(d) Provision of emergency cards

In the United Kingdom there has recently been interest in providing suicide attempters with cards which indicate how they might get emergency help at times of crisis. Two initial, relatively small, studies of this approach, one involving adults and the other young adolescents, produced encouraging results but a larger evaluation did not show the cards to be effective.[20] Provision of emergency cards requires there to be a 24 h service to deal with emergency calls. They might be thought helpful in a minority of cases, but there needs to be careful selection of patients who are offered this facility because of risk of it possibly being abused.

Pharmacological treatments

There have been relatively few treatment trials evaluating the effectiveness of pharmacological agents in suicide attempters. This perhaps reflects the problems of compliance with therapy, which were noted earlier, and risk of overdose.

(a) Antidepressants

A trial in the Netherlands in which paroxetine was compared with placebo in patients who were all repeaters of self-harm but who did not suffer from current depressive disorder showed apparent benefits for a subgroup of patients who received paroxetine, namely

those who had a history of between one and four episodes of self-harm. Patients with a history of five or more episodes did not seem to benefit.[21] The findings of this study are clearly of interest (although *post hoc* subgroup analyses of this kind must be treated with caution). Recently there has been much attention to the risk of antidepressants increasing suicidal ideation and acts in adolescents.[22] Also, it has become clear that there is increased risk with all types of antidepressants during the initial period of treatment.[23] These findings have highlighted the need to be cautious in the use of antidepressants, to provide early follow-up after initiating therapy, and to consider combining antidepressant treatment with other therapies, especially for adolescents (in whom only fluoxetine is currently recommended for the treatment of depression).

(b) Neuroleptics

A trial in which the depot neuroleptic flupenthixol was administered monthly in a dose of 20 mg for 6 months to repeaters of self-harm and compared with placebo in similar patients appeared to show that the active drug was effective in reducing the recurrence of self-harm.[24] While this type of study requires replication, perhaps using one of the atypical oral neuroleptics in patients who frequently repeat self-harm may be worth trying.

(c) Lithium

A systematic review of trials of lithium therapy versus a range of other drugs and placebo in patients with affective disorders has shown convincing evidence that lithium may prevent suicide.[25] It is not known if it may be anti-suicidal in other groups of patients.

Management in clinical practice

Before a treatment plan can be formulated a careful assessment must be carried out. In conducting the assessment the clinician needs to try and establish good rapport with the patient and be sensitive to the patient's preferences. The key factors that should be covered during the assessment are listed in Table 4.15.4.2. For the purpose of formulating a management plan it is particularly useful to draw up a problem list which summarizes the patient's current difficulties. This should be done in active collaboration with the patient as far as possible. Qualitative studies have shown that patients appreciate clinicians and other staff keeping them well informed of their mental and physical status and including them in decision-making in regard to their care.[26]

(a) Assessment

During the assessment it is crucial to estimate the risk of suicide or another non-fatal attempt. However, accurate assessment is far from easy. Risk factors for suicide following attempted suicide are shown in Table 4.15.4.3. Because suicide is uncommon, the predictive value of the items is limited. One predictor of suicide risk is the degree of suicidal intent involved in the current attempt (see Table 4.15.4.4). Clinicians should consider the use of the valuable Beck Suicide Intent scale.[27] Factors known to be associated with risk of a further attempt are listed in Table 4.15.4.5. It should be noted that while individuals who score positive on several of these factors will have considerably increased risk of repetition, a substantial proportion of those who repeat will not show these characteristics, i.e. the predictive value of scales to predict repetition is modest.

Table 4.15.4.2 The assessment of attempted-suicide patients

Factors that should be covered
Life events that preceded the attempt
Motives for the act, including suicidal intent and other reasons
Problems faced by the patient
Psychiatric disorder
Personality traits and disorder
Alcohol and drug misuse
Family and personal history
Current circumstances
Social (e.g. extent of social relationships)
Domestic (e.g. living alone or with others)
Occupation (e.g. whether employed)
Psychiatric history, including previous suicide attempts
Assessments that should be made
Risk of a further attempt
Risk of suicide
Coping resources and supports
What treatment is appropriate to the patient's needs
Motivation of patient (and significant others where appropriate, to engage in treatment)

Table 4.15.4.3 Factors associated with risk of suicide after attempted suicide

Older age
Male gender
Unemployed or retired
Separated, divorced or widowed
Living alone
Poor physical health
Psychiatric disorder (particularly depression, alcoholism, schizophrenia, and 'sociopathic' personality disorder)
High suicidal intent in current episode
Violent method involved in current attempt (e.g. attempted hanging, shooting, jumping)
Leaving a note
Previous attempt(s) (including repetitive self-injury)

Table 4.15.4.4 Factors that suggest high suicidal intent

Act carried out in isolation
Act timed so that intervention unlikely
Precautions taken to avoid discovery
Preparations made in anticipation of death (e.g. making will, organizing insurance)
Preparations made for the act (e.g. purchasing means, saving up tablets)
Communicating intent to others beforehand
Extensive premeditation
Leaving a note
Note alerting potential helpers after the act
Subsequent admission of suicidal intent

Table 4.15.4.5 Factors associated with risk of repetition of attempted suicide

Previous attempt(s)
Depression
Personality disorder
Alcohol or drug abuse
Previous psychiatric treatment
Unemployment
Lower socio-economic status
Criminal record
History of violence
Age 25–54 years
Single, divorced, or separated

(b) Treatment

The treatment plan should be drawn up on the basis of the patient's needs and risks. Inpatient psychiatric treatment will usually be indicated for patients with severe psychiatric disorders, especially where immediate risk of suicide appears to be high.

(i) Psychiatric disorders

Major psychiatric disorders should be treated in the usual way, but with particular care about use of medication which might be toxic in overdose. Specific treatment should be provided for alcohol and drug abuse; indeed, a suicide attempt is sometimes the first occasion that abuse may come to clinical attention.

(ii) Community therapy

Most patients can be managed in the community. Brief psychological therapy, with a focus on problem-solving will be appropriate for those patients who have clear problems, such as in interpersonal relationships, employment, or social adjustment. Some form of outreach (e.g. home visiting, telephone contact) may be helpful to increase the proportion of patients who engage in treatment, but is not necessary for most patients. Outreach may be essential in the treatment of patients in remote rural areas in developing countries.

If possible there should be continuity of therapy in terms of the same person who saw the patient in hospital after their attempt providing aftercare as this is likely to result in better compliance with therapy. Longer term cognitive behavioural therapy or dynamic psychotherapy may be required for patients whose attempts are related to traumas, such as sexual abuse, or to personality disorder. People who are repeaters of suicide attempts may also require more intensive treatment, especially those who frequently repeat. If resources permit, the use of a programme based on dialectical behaviour therapy, possibly using a group format for at least part of treatment, might be considered.

(iii) Adolescents

Family therapy may be required for young adolescents, and also for patients with difficulties in relation to children. Group therapy may be helpful for adolescents who are repeaters of self-harm.[28]

Prevention of suicide and attempted suicide

A widely diverse group of individuals are at risk of suicidal behaviour and it occurs in relation to a wide range of problems and situations. For example, suicide may occur in the context of long-term difficulties extending back to childhood, acute severe life events or, and perhaps most importantly, acute or long-term and relapsing mental illness. Because of this the range of potential prevention strategies is also considerable.

Suicide prevention programmes have been established in many countries. This is to be welcomed, not only because of the potential benefits in terms of suicide prevention, but also because of the likely benefits for the broader population of individuals with mental health problems. When considering prevention strategies, it is important to be aware of and sensitive towards issues relating to culture and ethnicity. For example, while suicide rates are generally relatively low in young females in the United Kingdom, this is not the case in young Asian females of the Hindu faith, in which rates appear to be relatively high and greater than those of their male peers.[29] The issues surrounding such deaths are often related to cultural clashes regarding values and expectations between young Asian females and their parents.

While ethical issues in relation to suicide prevention are not dealt with in detail in this chapter, they are none the less highly important.[30] Opinions will vary, for example, about whether suicide should always be prevented. This particularly relates to suicides occurring in the context of terminal and/or painful physical illnesses, and relapsing and debilitating mental illnesses. The ethics of suicide prevention overlap those of assisted suicide and euthanasia. Psychiatrists are increasingly likely to be drawn into debate and controversy about the ethical aspects of these issues, particularly in relation to severe and chronic mental illness, and mental health aspects of assisted suicide and euthanasia in people with severe physical illnesses.

General principles of prevention

Broadly there are two approaches to suicide prevention.[31,32] As described by Rose[33] in the context of prevention of health problems in general, one can distinguish between population approaches, which aim to decrease risk in the population as a whole, and high-risk group strategies, in which specific groups that are at increased risk are targeted. High-risk group strategies often appear more attractive and realistic. However, risk factors for many disorders are widely spread in the population and so the high-risk strategy tends to exclude a large number of people at moderate risk and is often ineffective in reducing the burden of a disease at the population level. Conversely, population strategies may appear more difficult to achieve but are more likely to be effective in reducing population levels of disease (see also Chapter 7.4). The main population and high-risk group strategies in the prevention of suicide and attempted suicide which are considered here are shown in Table 4.15.4.6.

It is unclear if national suicide prevention programmes are effective, although evidence of effectiveness for specific components of such strategies is emerging.[34] The most impressive programme, developed in Finland, was based on information from a detailed national study of all suicides in 1 year and includes a wide range of elements.[35] A decline in the Finnish suicide rate has been attributed to the programme.[36] In England a national suicide prevention strategy with a suicide target was introduced in 2002.[37] Strategies have also been introduced in Scotland[38] and Ireland.[39] While prevention strategies are difficult to evaluate[40] there are indications that programmes for prevention of suicide on a national scale may be effective.

Table 4.15.4.6 Examples of strategies for prevention of suicide and attempted suicide

Population strategies
Reducing availability of means for suicide
Educating primary care physicians
Influencing media portrayal of suicide
Educating the public about mental illness and its treatment
Educational approaches in schools
Befriending agencies and telephone helplines
Addressing the economic factors associated with suicidal behaviour
High-risk strategies
Prevention of suicide in:
Patients with psychiatric disorders
The elderly
Suicide attempters
High-risk occupational groups
Prisoners

Population strategies

(a) Reducing availability of means for suicide

This is the most widely discussed population strategy.[41] It is based on evidence that if the availability and/or danger of a popular method for suicide changes then this tends to have an impact on suicide rates. The general principles of prevention through reducing availability of means are, first, that many suicidal acts occur impulsively and therefore if a dangerous means is available this is more likely to result in death and, secondly, that the eventual suicide rate in survivors of serious attempts is remarkably low. Also the common adage that if people are intent on committing suicide they will find a means is not necessarily correct (see below).

(i) Coal gas

The most cited evidence for the effectiveness of this approach is the reduction in suicides in the United Kingdom which occurred in the 1960s and early 1970s when toxic coal gas supplies were gradually replaced with non-toxic North Sea gas.[42] Prior to this time coal gas poisoning through people placing their head in a gas oven was the most common method of suicide in the United Kingdom. As North Sea gas was gradually introduced the suicide rate dropped steadily, eventually being reduced by approximately a third. It is estimated that as many as 6000 deaths may have been prevented by this change. The effect also illustrates the point that when one method of suicide is no longer available people do not automatically turn to another, or if they do it may be to one that is less likely to cause death. Thus, it was some years before the suicide rate rose again, this being related to an increase in deaths from poisoning with carbon monoxide from car exhausts. Another factor that may have been relevant to the decline in suicides was the reduction in prescribing of barbiturates, these being replaced by far less toxic benzodiazepines.

(ii) Carbon monoxide

Suicide by carbon monoxide poisoning from car exhausts has become less common because cars are now fitted with catalytic converters. This has resulted in a decline in suicide rates in countries where this method of suicide had become more common, particularly in young males.[43]

(iii) Firearms

The widespread availability of guns in certain countries, particularly the United States, has been proposed as an important reason for their relatively high suicide rates. Guns are used in more than half of all suicides in the United States and their use for suicide correlates with the holding of gun licences in households.[44] Some controversy surrounds the question of whether restricting availability of guns leads to a reduction in suicide rates, but the weight of evidence seems to indicate that it does.[45]

(iv) Antidepressants

Given the very strong link between suicide and depression, and the risk of death from overdose of some of the older antidepressants, there has been much debate about whether more extensive use of newer, less toxic antidepressants would prevent suicides. This is not a simple question, as some patients respond better to the older tricyclic antidepressants. Another consideration concerns the cost of the newer antidepressants compared with the older varieties. Also it is very important to remember that most people who are taking antidepressants do not kill themselves with their antidepressants but use other methods. This and the probable selective prescribing of SSRIs to people judged to be at risk may account for the finding that suicide rates were higher in patients taking fluoxetine than patients taking other and in some cases more toxic antidepressants.[46] Nevertheless, common sense dictates that patients known to be at risk, and especially those with a history of suicidal behaviour, should be prescribed the less toxic preparations.

(v) Analgesics

In the United Kingdom and some other countries there has been particular concern about deaths from self-poisoning with paracetamol. Due to evidence that countries which have fewer tablets per pack seem to have a lower rate of mortality from paracetamol self-poisoning and because overdoses of paracetamol are often taken impulsively and involve household supplies, legislation was introduced in the United Kingdom in 1998 to reduce in the number of tablets of paracetamol (and aspirin) available per pack. This resulted in fewer overdoses, decreased cases of hepatotoxicity due to paracetamol toxicity, and a reduced number of deaths from both paracetamol and aspirin.[47]

(vi) Safety measures

Much attention has been paid to improving safety at popular sites for suicide. This includes erecting suicide barriers on bridges, multi-storey car parks, and other sites. If environmental changes are made such that a popular suicide site becomes safer, this does not mean that people at risk automatically move to using another site. For example, erection of barriers on the Clifton Suspension Bridge in Bristol, a popular site for suicide, has resulted in far fewer deaths by jumping.[48]

Clinicians involved in the development of suicide prevention strategies should look very carefully for local patterns which might provide clues about potentially effective measures for reducing access to methods. This could include, for example, ensuring that psychiatric inpatient units are free of hooks, pipes, and other objects or structures from which patients could hang themselves, and that all bed rails are collapsible (compulsory in the United Kingdom), secure fencing of railway lines or waterways close to psychiatric hospitals, and making local popular sites for suicide safer (e.g. suicide barriers on bridges). In addition, attention should

be paid to common dangerous methods of self-poisoning. Specific strategies may be required depending on local patterns. For example, the high rates of suicide in rural areas of developing countries due to self-poisoning with pesticides might be reversed with safe-storage programmes.

(b) Education of primary care physicians

Much of the attention regarding improved detection of individuals at risk has concerned the detection and management of depression in general practice. This was stimulated by findings that showed many people who died by suicide or who attempted suicide had seen their general practitioners shortly before these acts. Evidence that an intensive educational programme for general practitioners might be effective in influencing suicide rates comes from a study conducted on the Swedish island of Gotland.[49] In the year following this programme the suicide rate dropped significantly, prescribing of antidepressants by general practitioners increased, referrals to psychiatry, especially for depression, decreased, the amount of time lost from work for depression decreased, as did psychiatric admissions. Unfortunately this effect was fairly short-lived in that suicide rates rose again in subsequent years, which the authors attributed to some of the general practitioners having left the island. They also suggested that such programmes need to be repeated.[50] It is also important to note that the suicide rate only declined in females. The evidence in this study was based on relatively small numbers, at least with reference to suicide, although the effects on the management and outcome of depression were perhaps more impressive.

While the Gotland study has generated a lot of debate about suicide prevention in primary care, detection of people most at risk in general practice is extremely difficult because a large number of patients share risk factors and because suicide is a rare event. The most pragmatic view is that effective detection and treatment of depression (and other psychiatric disorders) in primary care are extremely important aims in their own rights and that they might also have benefits in terms of preventing some suicides.

Psychiatrists involved in designing suicide prevention strategies might ensure that there are effective local educational programmes for clinicians in primary care and other settings regarding detection and treatment of people with mental disorders.

(c) Influencing media portrayal of suicidal behaviour

Dramatic reporting and portrayal of suicidal behaviour by the media can facilitate suicidal acts in other people. This has been shown in a wide range of studies of both newspaper and television reporting of suicides and fictional presentations of suicidal behaviours in films and television dramas. The impact of media presentations appears to be greatest where the method used in the suicidal act is described in detail, where details of the deceased and/or the site of the act is provided, and for deaths by suicide of celebrities. The largest impact of media influence is on young people, although there is also influence on older people.[51]

In each country, consideration should be paid to the development of consensus statements about media policies in relation to reporting and portrayal of suicide,[52] which could be produced by joint working parties including representatives of the press, clinical and voluntary agencies, and experts in the field of suicidal behaviour. More difficult is the potentially valuable task of encouraging a policy whereby the media can be used to portray effective coping strategies for people in distress. Such a strategy will need to encompass local cultural factors. Psychiatrists and other professionals developing suicide prevention strategies might examine the practices of their local media with regard to reporting of suicides and, if necessary, hold meetings with media producers to explain the dangers of dramatic and extensive reporting, and also to explore how the media might help in prevention.

Concern is growing about influences on suicidal behaviour through the internet, especially web sites providing instructions on methods of suicide and chat rooms whereby individuals can instruct. Some sites seem to be intended to promote suicide, such as those which initiate meetings between suicidal individuals. Attempts can be made to regulate such sites, but this impossible for sites from other countries. Also, internet providers can be encouraged to ensure that sites offering positive help appear before less desirable sites.

(d) Education of the public about mental illness and its treatment

In view of the very strong link between suicide and mental illness, effective treatment of psychiatric disorders must be a central theme in suicide prevention. However, detection of people with disorders will depend on the awareness that they and those around them have regarding the signs and symptoms of disorder, and their willingness to seek appropriate help.[53] These important stages in receiving effective help will depend on the general public's attitudes towards mental illness and knowledge of its nature and the feasibility of treatment. In several countries, programmes to encourage education of the public about psychiatric disorders (especially depression) and to tackle stigmatization of those who are ill have been established. At this stage, evidence is lacking as to whether or not they have been successful. Psychiatrists and their colleagues might consider similar campaigns where these are not already in place, although the method of delivery of messages (e.g. media presentations, leaflets, workshops, articles) will clearly depend on local factors.

(e) Educational approaches in schools

There have been three broad approaches in trying to prevent suicide through school-based programmes.[54] The first of these includes teaching about the facts of suicide. Worrying evidence from the United States that such a programme appeared to lead to a small increase in pupils' ratings of the acceptability of suicide as an option compared with the ratings of pupils who did not receive the programme suggests that this is not a wise approach.

Suicidal behaviour in young people often appears to be related to depression, anxiety, low self-esteem, difficulties during upbringing (e.g. abuse, deprivation), life events (especially break-up of relationships, family problems, and bullying), and poor problem-solving skills.[54] Also, troubled and suicidal young people most often seek help from their peers. A second school-based strategy has been the development of educational programmes in schools about recognition of psychological distress in individuals and their peers, problem-solving, and peer support. Given the early age at which suicidal behaviour begins, such programmes should probably be targeted at extremely young school children, with later sessions for adolescents.

A third approach is to screen adolescents with questionnaires to detect children and adolescents at risk of psychiatric disorder and

possible suicidal behaviour. Pupils that are so detected will then need referral to an appropriate agency for further assessment and possible treatment. While there is some evidence to support such an approach it is not without drawbacks. (Suicide in children and adolescents is considered further in Chapter 9.2.10).

For psychiatrists and others involved in developing local prevention strategies it is important to recognize that school-based approaches to prevention represents is a highly sensitive area and one where the most effective (and least risky) approach is at present unclear. Another important aspect of suicidal behaviour in school pupils is the management of the aftermath of suicides and its impact on other pupils and how to tackle outbreaks of self-harm.[54]

(f) Befriending agencies and telephone helplines

A very important component of suicide prevention policy in many countries is the support provided by largely volunteer staffed befriending agencies and especially telephone helplines. The best known of these is Samaritans. A key principle on which such services are based is that people in distress and at risk of suicide will benefit from being able to discuss their problems with someone entirely confidentially. Recently, more assertive outreach programmes, in which volunteers meet up with distressed individuals such as in prisons and in remote areas, have been added to the traditional telephone service. In the United Kingdom and elsewhere counselling by e-mail and text messaging is being used extensively.

The effectiveness of these approaches is largely unknown. Conducting controlled trials to examine their efficacy is very difficult. Naturalistic studies have produced conflicting evidence about the effectiveness of the Samaritans in the United Kingdom.[55,56] An examination of changes in suicide rates in areas with and without crisis intervention services in the United States suggested that suicide rates in young white females may have been reduced in areas where such services were developed.[57] Given the large numbers of contacts made with Samaritans in the United Kingdom (nearly 5 million in 2005), it appears that the service is valued by people in distress.

Volunteer-run telephone helplines and similar services may benefit greatly from the support and advice of local clinicians, who should regard them as a potentially valuable element in a local suicide prevention strategy.

Addressing the economic factors associated with suicidal behaviour

The association between suicidal behaviour and unemployment and poverty suggests that in order for suicide rates to change markedly these important socio-economic factors must be modified. The big increase in suicide rates during the economic depression of the late 1920s and early 1930s and the statistical association between suicide risk and unemployment would support this. Clearly such factors are increasingly a reflection of the global economic situation, but the strategies of individual governments, particularly in relation to the employment prospects for young people, may be influential. The main role of psychiatrists may be in highlighting these factors. The considerable evidence that changes in the economic environment can exert a powerful influence on suicide rates indicates that governments with serious intentions to reduce suicide rates should address these issues.[32]

Strategies for high-risk groups

There are a wide range of possible prevention strategies which can be targeted at high-risk groups. Here are some of the more important examples of such groups and relevant strategies will be discussed.

(a) Patients with psychiatric disorders

(i) Risk identification

One approach to preventing suicide in people with known psychiatric disorders is to try and use recognized risk factors for suicide in each disorder to identify high-risk patients. The main psychiatric disorders in series of people who have died by suicide are depression (approximately two-thirds), severe alcohol abuse (approximately 15 per cent), and schizophrenia (5–10 per cent).[58] The main suicide risk factors identified in these three disorders include, for example, previous attempts, family history of suicidal behaviour, and living alone. Comorbidity of disorders (e.g. depression and alcohol abuse) and of personality and psychiatric disorders increases risk.

One difficulty in using a risk-identification approach, however, is that the risk factors identified from studies of groups of individuals who have died from suicide are often misleading when applied to individual patients. Also when applying such factors a relatively large number of individuals will appear to be at risk when they may not in fact be so. In clinical practice it is important to be aware of patients who, because of their individual characteristics, are at long-term high risk. Clinicians must also be aware of acute situations which may temporarily increase the risk in patients, be they ones at long-term risk or not.

The most pragmatic approach, therefore, is to ensure that proven effective treatments for patients with these conditions are available and also to be particularly cautious at times of obvious high risk. There are particular periods of risk of suicide for patients with psychiatric disorders. One of these is during the first few weeks after discharge from psychiatric hospital.[59] This emphasizes the necessity for continuity of care at this critical time. Other risk times may be following the break-up of a relationship or other significant loss, during periods of marked hopelessness, shortly after discharge from hospital, and following recent suicidal behaviour by another patient or someone else close to the individual.

(ii) Preventative strategies

Prevention of suicide in patients with psychiatric disorders must be a major element in any suicide prevention strategy.[60] Important strategies in preventing suicide in patients with affective disorders include active treatment of individual episodes of illness, psychological therapy to improve compliance with treatment and assisting individuals to manage their disorder, use of lithium and other mood stabilizers for patients with recurrent bipolar disorders, and use of long-term antidepressants in patients with frequent relapses of depressive disorders. Risk is often greatest during the early stages of a disorder.[53]

The risk factors in schizophrenia indicate that risk tends to be highest between episodes of acute illness when patients may have insight and feel hopeless about their circumstances and prospects. Risk is related more to affective symptoms than core features of the disorders.[61] Continuity of care is likely to be a particularly important factor in preventing suicides in such patients at risk, with care

being continued energetically during periods of remission. Community psychiatric nurses have a very important role with such patients. The use of the newer atypical neuroleptics may also be beneficial.[62]

Direct treatment of abuse is likely to be the best preventive strategy for patients with substance abuse disorders, with care taken to manage episodes of depression. The particularly high risk in the weeks following a break-up of a relationship for patients with severe alcohol abuse[63] again points to the need for continuity of support in the community.

The prevention of suicide in patients with comorbid disorders, especially the combination of depression with alcohol abuse and/or personality disorder, is a challenging task, particularly as compliance with treatment is often less good than in patients with single disorders. Effective prevention is likely to depend on close integration of care between different statutory care agencies.

Another important element in prevention in this population is education in suicide risk assessment and management procedures for clinical staff at all levels of seniority. These should be incorporated in educational programmes for risk assessment in general.

(b) Elderly people

In planning suicide prevention in the elderly population, account must be taken of the relative immobility of many older people. In a region of Italy, introduction of a telephone service to provide support and access to emergency help for elderly persons at risk has been associated with an encouraging decline in elderly suicides in the area.[64] This might serve as a model for other countries.

(c) Suicide attempters

In view of the clear association between non-fatal suicidal behaviour and subsequent suicide, establishment of adequate services for suicide attempters, including the provision of careful assessments of patients in the general hospital and offering treatments for which at least some indicators of benefit are available (see above), is an important element in any national suicide prevention strategy. There is good evidence that well-trained, non-medical psychiatric staff can effectively carry out assessments and arrange aftercare. Models for ideal services exist, such as those published by the National Institute for Clinical Excellence[65] in the United Kingdom.

(d) High-risk occupational groups

Certain occupational groups are known to be at relatively high risk of suicide. In the United Kingdom these include farmers, veterinary surgeons, dental practitioners, medical practitioners, pharmacists, and female nurses. It is interesting to note that all these groups have relatively easy access to dangerous methods for suicide. Since prevention through detection of those most at risk encounters the usual difficulties of prevention of relatively rare behaviour using rather crude risk factors, it is probably more important to have general strategies for improving care in individual groups. In doctors, for example, there are some particular difficulties about confidentiality and therefore providing easy means of doctors getting confidential help is important. In farmers, improving the knowledge and attitudes of farming communities towards psychiatric disorder, and removing access to firearms at times of risk, are likely to be important.

(e) Prisoners

There are relatively high suicide rates in prisoners,[66] especially young males held on remand. While one aspect of prevention is through ensuring that prisons and police cells are safe in terms of absence of structures from which inmates can hang themselves, there are a range of other potentially useful and humane strategies. These include careful assessments of new inmates using risk-assessment procedures, training of staff with regard to both assessment skills and attitudes towards mental health problems and suicide prevention, in-reach programmes by befriending organizations such as the Samaritans, and ready access to psychiatric and psychological services. Clinicians involved in local suicide prevention programmes should include prisons in their considerations.

Further information

Hawton, K. (2005). *Prevention and treatment of suicidal behaviour: from science to practice*. Oxford University Press, Oxford.

Hawton, K. and Van Heeringen, K. (2000). *The international handbook of suicide and attempted suicide*. Wiley, Chichester.

National Collaborating Centre for Mental Health. (2004). *Self-harm: the short-term physical and psychological management and secondary prevention of self-harm in primary and secondary care (full guideline) clinical guideline 16*. National Institute for Clinical Excellence, London.

References

1. World Health Organization. (1994). *Ninth general programme of work covering the period 1996–2000*. World Health Organization, Geneva.
2. United Nations. (1996). *Prevention of suicide: guidelines for the formulation and implementation of national strategies*. Department for Policy Coordination and Sustainable Development, New York.
3. Hjelmeland, H., Hawton, K., Nordvik, H., *et al.* (2002). Why people engage in parasuicide: a cross-cultural study of intentions. *Suicide & Life Threatening Behavior*, **32**, 380–93.
4. Zahl, D. and Hawton, K. (2004). Repetition of deliberate self-harm and subsequent suicide risk: long-term follow-up study in 11,583 patients. *The British Journal of Psychiatry*, **185**, 70–5.
5. Owens, D., Horrocks, J., and House, A. (2002). Fatal and non-fatal repetition of self-harm systematic review. *The British Journal of Psychiatry*, **181**, 193–9.
6. Haw, C., Hawton, K., Houston, K., *et al.* (2001). Psychiatric and personality disorders in deliberate self-harm patients. *The British Journal of Psychiatry*, **178**, 48–54.
7. Paykel, E.S., Prusoff, B.A., and Myers, J.K. (1975). Suicide attempts and recent life events: a controlled comparison. *Archives of General Psychiatry*, **32**, 327–33.
8. Williams, J.M.G., Crane, C., Barnhofer, T., *et al.* (2005). Psychology and suicidal behaviour: elaborating the entrapment model. In *Prevention and treatment of suicidal behaviour: from science to practice* (ed. K. Hawton), pp. 71–90. Oxford University Press, Oxford.
9. Mann, J.J. (2003). Neurobiology of suicidal behaviour. *Nature Reviews Neuroscience*, **4**, 819–28.
10. Beck, A.T., Steer, R.A., Kovacs, M., *et al.* (1985). Hopelessness and eventual suicide: a 10 year prospective study of patients hospitalised with suicidal ideation. *The American Journal of Psychiatry*, **145**, 559–63.
11. Hawton, K., Townsend, E., Arensman, E., *et al.* (2005). Psychosocial and pharmacological treatments for deliberate self harm *Cochrane Database of Systematic Reviews*, Issue: 4, Art. No.: CD001764. DOI: 10.1002/14651858.CD001764.

12. Townsend, E., Hawton, K., Altman, D.G., *et al.* (2001). The efficacy of problem-solving treatments after deliberate self-harm: meta-analysis of randomised controlled trials with respect to depression, hopelessness and improvement in problems. *Psychological Medicine*, **31**, 979–88.
13. Brown, G.K., Have, T.T., Henriques, G.R., *et al.* (2005). Cognitive therapy for the prevention of suicide attempts: a randomized controlled trial. *The Journal of the American Medical Association*, **294**, 563–70.
14. Tyrer, P., Thompson, S., Schmidt, U., *et al.* (2003). Randomized controlled trial of brief cognitive behaviour therapy versus treatment as usual in recurrent deliberate self-harm: the POPMACT study. *Psychological Medicine*, **33**, 969–76.
15. Linehan, M.M., Armstrong, H.E., Suarez, A., *et al.* (1991). Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Archives of General Psychiatry*, **48**, 1060–4.
16. Linehan, M.M., Comtois, K.A., Murray, A.M., *et al.* (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*, **63**, 757–66.
17. Van Heeringen, C., Jannes, S., Buylaert, W., *et al.* (1995). The management of non-compliance with referral to out-patient after-care among attempted suicide patients: a controlled intervention study. *Psychological Medicine*, **25**, 963–70.
18. Vaiva, G., Ducrocq, F., Meyer, P., *et al.* (2006). Effect of telephone contact on further suicide attempts in patients discharged from an emergency department: randomised controlled study. *British Medical Journal*, **332**, 1241–5.
19. Carter, G.L., Clover, K., Whyte, I.M., *et al.* (2005). Postcards from the EDge project: randomised controlled trial of an intervention using postcards to reduce repetition of hospital treated deliberate self poisoning. *British Medical Journal*, **331**, 805–9.
20. Evans, J., Evans, M., Morgan, H.G., *et al.* (2005). Crisis card following self-harm: 12-month follow-up of a randomised controlled trial. *The British Journal of Psychiatry*, **187**, 186–7.
21. Verkes, R.J., Van-der-Mast, R.C., Hengeveld, M.W., *et al.* (1998). Reduction by paroxetine of suicidal behavior in patients with repeated suicide attempts but not major depression. *The American Journal of Psychiatry*, **155**, 543–7.
22. Committee on Safety of Medicines. (2003). Selective serotonin reuptake inhibitors (SSRIs): overview of regulatory status and CSM advice relating to major depressive disorder (MDD) in children and adolescents including a summary of available safety and efficacy data. Medicines and Healthcare products Regulatory Agency, London.
23. Jick, H., Kaye, J.A., and Jick, S.S. (2004). Antidepressants and the risk of suicidal behaviors. *The Journal of the American Medical Association*, **292**, 338–43.
24. Montgomery, S.A., Montgomery, D.B., Jayanthi-Rani, S., *et al.* (1979). Maintenance therapy in repeat suicidal behaviour: a placebo controlled trial. Proceedings of the 10th International Congress for Suicide Prevention & Crisis Intervention. International Association for Suicide Prevention, pp. 227–9. Ottawa, Canada.
25. Cipriani, A., Pretty, H., Hawton, K., *et al.* (2005). Lithium in the prevention of suicidal behaviour and all-cause mortality in patients with mood disorders: a systematic review of randomised trials. *The American Journal of Psychiatry*, **162**, 1805–19.
26. Horrocks, J., Hughes, J., Martin, C., *et al.* (2005). *Patient experiences of hospital care following self-harm—a qualitative study*. Academic Unit of Psychiatry and Behavioural Sciences, Leeds.
27. Beck, A., Schuyler, D., and Herman, J. (1974). Development of suicidal intent scales. In *The prediction of suicide* (eds. A. Beck, H. Resnik, and D.J. Lettieri), pp. 45–56. Charles Press, Maryland.
28. Wood, A., Trainor, G., Rothwell, J., *et al.* (2001). Randomized trial of group therapy for repeated deliberate self-harm in adolescents. *Journal of American Academy of Child and Adolescent Psychiatry*, **40**, 1246–53.
29. Soni Raleigh, V. and Balarajan, R. (1992). Suicide and self-burning among Indians and West Indians in England and Wales. *The British Journal Psychiatry*, **161**, 365–8.
30. Burgess, S. and Hawton, K. (1998). Suicide, euthanasia and the psychiatrist. *Philosophy, Psychiatry and Psychology*, **5**, 113–26.
31. Hawton, K. (2005). *Prevention and treatment of suicidal behaviour: from science to practice*. Oxford University Press, Oxford.
32. Lewis, G., Hawton, K., and Jones, P. (1997). Strategies for preventing suicide. *The British Journal Psychiatry*, **171**, 351–4.
33. Rose, G. (1992). *The strategy of preventive medicine*. Oxford University Press, Oxford.
34. Mann, J.J., Apter, A., Bertolote, J., *et al.* (2005). Suicide prevention strategies. A systematic review. *The Journal of the American Medical Association*, **294**, 2064–74.
35. Upanne, M. (1998). Implementation of the suicide prevention strategy in Finland: first follow-up. In *Suicide prevention: a holistic approach* (eds. D. De Leo, A. Schmidtke, and R.F.W. Diekstra), pp. 219–23. Kluwer Academic Publishers, Dordrecht.
36. Ohberg, A. and Lönnqvist, J. (1997). Suicide trends in Finland 1980–1995. *Psychiatrica Fennica*, **28**, 11–23.
37. Department of Health. (2002). *National suicide prevention strategy for England*. Department of Health, London.
38. Scottish Executive. (2002). Choose life: a national strategy and action plan to prevent suicide in Scotland. The Stationery Office, Edinburgh.
39. Health Service Executive, the National Suicide Review Group, Department of Health and Children. (2005). *Reach out: national strategy for action on suicide prevention 2005–2014*. Health Service Executive, Dublin.
40. Goldney, R.D. (1998). Suicide prevention is possible: a review of recent studies. *Archives of Suicide Research*, **4**, 329–39.
41. Hawton, K. (2005). Restriction of access to methods of suicide as a means of suicide prevention. In *Prevention and treatment of suicidal behaviour: from science to practice* (ed. K. Hawton), pp. 279–91. Oxford University Press, Oxford.
42. Kreitman, N. (1976). The coal gas story: United Kingdom suicide rates 1960–1971. *British Journal of Preventive and Social Medicine*, **30**, 86–93.
43. Amos, T., Appleby, L., and Kiernan, K. (2001). Changes in rates of suicide by car exhaust asphyxiation in England and Wales. *Psychological Medicine*, **31**, 935–9.
44. Kellermann, A.L., Rivara, F.P., Somes, G., *et al.* (1992). Suicide in the home in relation to gun ownership. *New England Journal of Medicine*, **327**, 467–72.
45. Youth Suicide by Firearms Task Force. (1998). Consensus statement on youth suicide by firearms. *Archives of Suicide Research*, **4**, 89–94.
46. Jick, S.S., Dean, A.D., and Jick, H. (1995). Antidepressants and suicide. *British Medical Journal*, **310**, 215–8.
47. Hawton, K., Simkin, S., Deeks, J., *et al.* (2004). UK legislation on analgesic packs: before and after study of long term effect on poisonings. *British Medical Journal*, **329**, 1076–9.
48. Bennewith, O., Nowers, M., and Gunnell, D. (2007). Effect of barriers on the Clifton suspension bridge, England, on local patterns of suicide: implications for prevention. *The British Journal Psychiatry*, **190**, 266–7.
49. Rutz, W., von Knorring, L., and Walinder, J. (1989). Frequency of suicide on Gotland after systematic postgraduate education of general practitioners. *Acta Psychiatrica Scandinavica*, **80**, 151–4.
50. Rutz, W., von Knorring, L., and Walinder, J. (1992). Long-term effects of an educational program for general practitioners given by the Swedish committee for the prevention and treatment of depression. *Acta Psychiatrica Scandinavica*, **85**, 83–8.
51. Hawton, K. and Williams, K. (2005). Media influences on suicidal behaviour: evidence and prevention. In *Prevention and treatment of suicidal behaviour: from science to practice* (ed. K. Hawton), pp. 293–306. Oxford University Press, Oxford.

52. Pirkis, J., Blood, R.W., Beautrais, A., *et al.* (2006). Media guidelines on the reporting of suicide. *Crisis*, **27**, 82–7.
53. Jamison, K.R. and Hawton, K. (2005). The burden of suicide and clinical suggestions for prevention. In *Prevention and treatment of suicidal behaviour: from science to practice* (ed. K. Hawton), pp. 183–96. Oxford University Press, Oxford.
54. Hawton, K. and Rodham, K. (2006). *By their own young hand: deliberate self-harm and suicidal ideas in adolescents*. Jessica Kingsley Publishers, London.
55. Bagley, C.R. (1968). The evaluation of a suicide prevention scheme by an ecological method. *Social Science & Medicine*, **2**, 1–14.
56. Jennings, C., Barraclough, B.M., and Moss, J.R. (1978). Have the Samaritans lowered the suicide rate? A controlled study. *Psychological Medicine*, **8**, 413–22.
57. Lester, D. (1974). Effect of suicide prevention centers on suicide rates in the United States. *Health Service Reports*, **89**, 37–9.
58. Cavanagh, J.T.O., Carson, A.J., Sharpe, M., *et al.* (2003). Psychological autopsy studies of suicide: a systematic review. *Psychological Medicine*, **33**, 395–405.
59. Goldacre, M., Seagroatt, V., and Hawton, K. (1993). Suicide after discharge from psychiatric inpatient care. *Lancet*, **342**, 283–6.
60. University of Manchester. (2006). Avoidable deaths. Five year report of the National Confidential Inquiry into suicide and homicide by people with mental illness. National Confidential Inquiry into Suicide and Homicide by People with Mental Illness, Manchester.
61. Hawton, K., Sutton, L., Haw, C., *et al.* (2005). Schizophrenia and suicide: a systematic review of risk factors. *The British Journal of Psychiatry*, **187**, 9–20.
62. Meltzer, H.Y., Alphs, L., Green, A.I., *et al.* (2003). Clozapine treatment for suicidality in schizophrenia. *Archives of General Psychiatry*, **60**, 82–91.
63. Murphy, G.E., Armstrong, J.W. Jr, Hermele, S.L., *et al.* (1979). Suicide and alcoholism: interpersonal loss confirmed as a predictor. *Archives of General Psychiatry*, **36**, 65–9.
64. De Leo, D., Buono, M.D., and Dwyer, J. (2002). Suicide among the elderly: the long-term impact of a telephone support and assessment intervention in northern Italy. *The British Journal Psychiatry*, **181**, 226–9.
65. National Institute for Clinical Excellence. (2004). Clinical guideline 16. Self-harm: the short-term physical and psychological management and secondary prevention of self-harm in primary and secondary care. National Institute for Clinical Excellence, London.
66. Fazel, S., Benning, R., and Danesh, J. (2005). Suicides in male prisoners in England and Wales, 1978–2003. *Lancet*, **366**, 1301–2.

4.16

Culture-related specific psychiatric syndromes

Wen-Shing Tseng

The concept of culture-related specific (psychiatric) syndromes

In certain ways, all psychiatric disorders are more or less influenced by cultural factors, in addition to biological and psychological factors, for their occurrence and manifestation. 'Major' psychiatric disorders (such as schizophrenia or bipolar disorders) are more determined by biological factors and relatively less by psychological and cultural factors, but 'minor' psychiatric disorders (such as anxiety disorders, conversion disorders, or adjustment disorders) are more subject to psychological causes as well as cultural factors. In addition to this, there are groups of psychiatric disorders that are heavily related to and influenced by cultural factors, and therefore addressed as culture-related specific psychiatric syndromes.

Culture-related specific psychiatric syndromes, also called culture-bound syndromes[1] or culture-specific disorders,[2] refer to mental conditions or psychiatric syndromes whose occurrence or manifestations are closely related to cultural factors and thus warrant understanding and management primarily from a cultural perspective. Because the presentation is usually unique, with special clinical manifestations, the disorder is called a culture-related specific psychiatric syndrome.[3] From a phenomenological point of view, such a condition is not easily categorized according to existing psychiatric classifications, which are based on clinical experiences of commonly observed psychiatric disorders in western societies, without adequate orientation towards less frequently encountered psychiatric conditions and diverse cultures worldwide.

Around the turn of the twentieth century, during a period of colonization by western societies, western ministers, physicians, and others visited faraway countries, where they encountered behaviours and unique psychiatric conditions that they had never experienced at home. Most of these conditions were known to the local people by folk names, such as *latah*, *amok*, *koro*, *susto*, and so on, and were described by westerners as exotic, rare, uncommon, extraordinary mental disorders, mental illnesses peculiar to certain cultures, or culture-bound syndromes. The latter term implies that such syndromes are bound to a particular cultural region.[4]

Recently, however, cultural psychiatrists have realized that such psychiatric manifestations are not necessarily confined to particular ethnic-cultural groups. For instance, epidemic occurrences of *koro* (penis-shrinking panic) occur among Thai or Indian people, not only among the Southern Chinese as originally claimed; and sporadic occurrences of *amok* attacks (mass, indiscriminate homicidal acts) are observed in the Philippines, Thailand, Papua New Guinea, and in epidemic proportions in many places in South Asia,[5] in addition to Malaysia where it was believed to most commonly occur. Terrifying examples of *amok* have recently occurred with frequency on school campuses and in workplaces in the United States.

Therefore, the term culture-bound does not seem to apply, and it has been suggested that culture-related specific psychiatric syndrome would be more accurate to describe a syndrome that is closely **related** to certain *cultural traits* or *cultural features* rather than **bound** specifically to any one *cultural system* or *cultural region*.[4] Accordingly, the definition has been modified to a collection of signs and symptoms that are restricted to a limited number of cultures, primarily by reason of certain of their psychosocial features,[6] even though it is recognized that every psychopathology is influenced by culture to a certain degree.

Subgrouping of culture-related specific syndromes

In order to organize and categorize the various culture-related syndromes, several different subgroup systems have been proposed by different scholars in the past, such as by cardinal symptoms or by 'taxon' (according to a common factor). However, instead of focusing on the clinical manifestation descriptively, it will be more meaningful to subgroup the syndromes according to how they might be affected by cultural factors.

It has been recognized that there are different ways culture contributes to psychopathology. Namely: *pathogenetic effect* (culture has causative effect), *pathoselective effect* (culture selects the nature and type of psychopathology), *pathoplastic effect* (culture contributes to the manifestation of psychopathology), *pathoelaborating effect* (culture elaborates and reinforces certain types of manifestations), *pathofacilitating effect* (culture contributes to the frequent occurrence of particular psychopathologies), or *pathoreactive effect* (culture determines the reaction to psychopathology). Furthermore, culture impacts differently on different types of psychopathology.

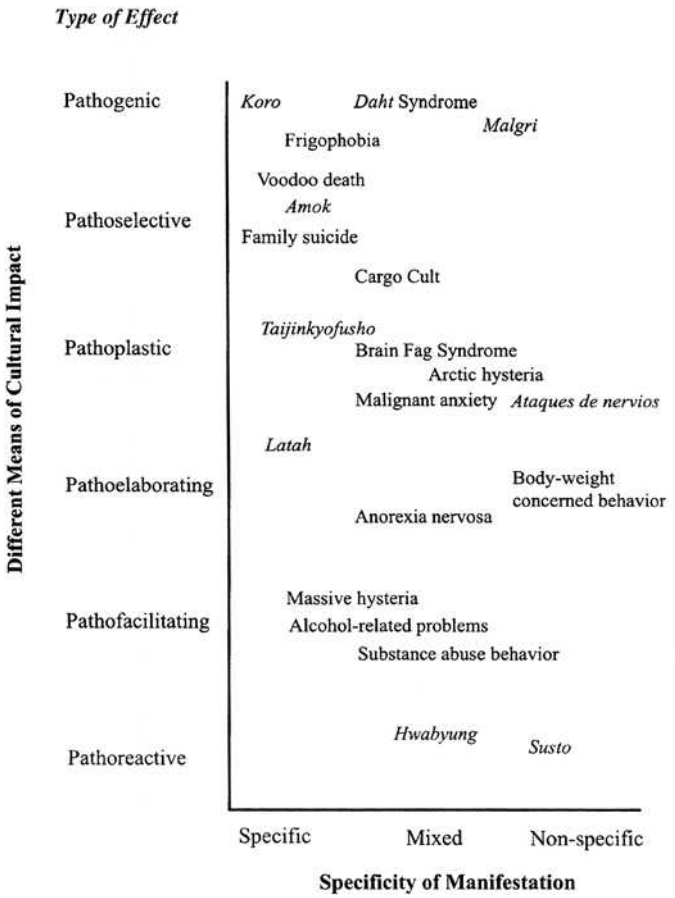


Fig. 4.16.1 Position of culture-related syndromes according to two parameters (This article was published in Handbook of Cultural Psychiatry, Tsend, W.S., copyright Elsevier (2001).)

If the psychopathology is divided into that which is more biologically determined (such as organic mental disorders or major psychiatric disorders), psychologically determined (such as minor psychiatric disorders), or socio-culturally determined (such as culture-related specific syndromes or epidemic mental disorders), it can be said that a pathoreactive effect is observed for all types of psychiatric disorders while a psychogenetic effect is observed mainly for those psychiatric disorders which are more socio-culturally determined, such as culture-related specific syndromes or epidemic mental disorders.

Academically identified culture-related specific syndromes can be compared according to two parameters, namely, the different means of cultural impact and the specificity of the manifestation. It will be useful to determine the basic ways culture contributes to the psychopathology, namely: pathogenic, pathoselective, pathoplastic, pathoelaborating, pathofacilitating, or pathoreactive effects; and to what extent the manifested syndrome is: specific, mixed, or non-specific (Fig. 4.16.1). For example, the primary cultural contribution to *koro* is a pathogenic effect with specific manifestation; to *latah* a pathoelaborating effect with specific manifestation; and to *susto* a pathoreactive effect with non-specific manifestation.

Culture-related beliefs as causes of the occurrence (pathogenetic effects)

(a) Koro (genital-retraction anxiety disorder)

(i) Definition

Koro is a Malay term which means the head of a turtle, symbolizing the male sexual organ which can 'shrink'. Clinically *koro* refers to the psychiatric condition in which the patient is morbidly concerned that his penis is shrinking excessively and subsequently dangerous consequences (such as death) might occur. The manifested symptoms may vary from simple excessive concern to obsessive/hypochondriac concern, intense anxiety, or a panic condition related to the shrinking of the penis. Clinically, this is usually a benign (non-psychotic) condition that occurs in individual, as a sporadic case, but occasionally it may grow to epidemic proportions, so that several hundreds or a thousand people may develop this disorder in a panic manner within a limited time of several weeks or several months. The majority of cases are young males who fear that their penises are shrinking. However, the organ concerned may be any protruding part of the body, such as the nose or ear (particularly when patients are prepuberty children) or the nipples or labia (in females).

Sporadic occurrences of female *koro* cases have never been reported. However, in *koro* epidemics, a small portion of the victims may be female.[7,8] In those cases, the female patients demonstrate slightly different clinical pictures, mainly focusing on the retraction of the nipples and some on the labia. The clinical condition is characterized by a more or less hysterical panic, associated with multiple somatic symptoms, a bewitched feeling, or the misinterpretation or accusation by others of being bewitched.

(ii) Geographic and ethnic group distribution

Cultural psychiatrists originally considered *koro* to be a culture-bound disorder related only to the Chinese. Most Chinese investigators have taken the view that the disorder is related to the Chinese cultural belief in *suoyang*, literally means shrinkage of yang organ due to excess loss of Yang (male) element. It has been speculated further that the occurrence of *koro* among people in South Asian countries, such as Malaysia and Indonesia, was the result of Chinese migrants. However, this cultural-diffusion view is doubted now, since *koro* epidemics have been reported in Thailand and India as well, involving non-Chinese victims.

As a result of the dissemination of knowledge about *koro* as a culture-bound syndrome, there is increased literature reporting so-called *koro* cases from various ethnic groups around the world, such as from the United Kingdom, Canada, and Israel. However, it was pointed out that among the cases reported, all suffered primarily from many psychiatric conditions: affective disorders, schizophrenia and anxiety disorders, drug abuse, or organic brain disorders. Therefore, they were referred as having *koro*-like symptoms, not exactly the same as the *koro* syndrome. It is necessary for clinicians to recognize that *koro* is referred to on different levels as: a symptom, a syndrome, or an epidemic disorder.

(iii) Epidemic of koro

Koro attack may occur occasionally as endemic or epidemic, involving many victims. Epidemic *koro* has been observed in several areas: Guangdong area of China, Singapore, Thailand, and India. As an epidemic, it is manifested as a panic state rather than an anxiety or obsessive state. The epidemic *koro* has tended to occur when there

was socio-political stress within the epidemic areas, and an outburst of *koro* by a group of people may be interpreted as a way to deal with the social stress that they encountered.

(iv) Diagnostic issue

Clinicians habitually try to fit pathologies into certain diagnostic categories. Because the patient, based on his misinterpretation, is morbidly preoccupied with the idea that certain ill-effects may occur due to the excessive retraction of his genital organ, the condition may, in a broad sense, be classified as a hypochondriacal disorder as defined in DSM-IV. The condition is also similar to a body dysmorphic disorder, as the patient is preoccupied with a culturally induced, imagined defect in his physical condition. If the focus is on how the patient reacts emotionally, how he responds to the culture-genic stress, with fear, anxiety, or a panic state, anxiety disorder may be considered. However, when we try to categorize culture-related specific syndromes according to the existing nosologically oriented classification system, their meaning and purpose are lost.

(v) Therapy

As for therapy of sporadic individual cases, assurance may be provided or medical knowledge offered in the form of educational counselling to eliminate the patient's concern about impending death. This supportive therapy may work in many cases, but, for someone who firmly believes the *koro* concept, it may not. In general, a young, unmarried male, who lacks adequate psychosexual knowledge and experience, will respond favourably to therapy. If necessary, it is desirable to work on issues such as the patient's self-image, self-confidence, or his masculinity.

(b) Dhat syndrome (semen-loss anxiety)

Very closely related to the genital-retraction anxiety disorder (*koro*) is the semen-loss or semen-leaking anxiety disorder, or spermatorrhea, also known by its Indian folk name, *dhat* syndrome. The *dhat* syndrome refers to the clinical condition in which the patient is morbidly preoccupied with the excessive loss of semen from an improper form of leaking, such as nocturnal emissions, masturbation, or urination. The underlying anxiety is based on the cultural belief that excessive semen loss will result in illness. Therefore, it is a pathogenically induced psychological disorder.[9,10] The medical term spermatorrhea is a misnomer, as there is no actual problem of sperm leakage from a urological point of view.

From a clinical point of view, the patients are predominantly young males who present vague, multiple somatic symptoms such as fatigue, weakness, anxiety, loss of appetite, and feelings of guilt (about having indulged in sexual acts such as masturbation or having sex with prostitutes). Some also complain of sexual dysfunction (impotence or premature ejaculation). The chief complaint is often that the urine is opaque, which is attributed to the presence of semen. The patient attributes the passing of semen in the urine to his excessive indulgence in masturbation or other socially defined sexual improprieties.[11] The syndrome is also widespread in Nepal, Sri Lanka (where it is referred to as *prameha* disease), Bangladesh, and Pakistan.

(c) Sorcery fear and voodoo death (magic-fear-induced death)

The peculiar phenomenon of voodoo death refers to the sudden occurrence of death associated with taboo-breaking or curse fear. It is based on the belief in witchcraft, the putative power to bring about misfortune, disability, and even death through spiritual mechanisms.[12] A severe fear reaction may result from such beliefs, which may actually end in death. From a psychosomatic point of view, it would be psychogenically induced death. From a cultural psychiatric perspective, it is another example of culture-induced morbid fear reaction.

Medically it has been recognized that sudden death was related to psychological stress occurring during experiences of acute grief, the threat of the loss of a close person, personal danger, or other stressful situations. It has been speculated that the cause of death was possibly from natural causes; the possible use of poisons due to sorcery or witchcraft; excessive fear reaction; or death from dehydration or existing physical illness due to old age.

Culture-patterned specific coping reactions (pathoselective effect)

(a) Amok (indiscriminate mass homicide attacks)

(i) The nature of the behaviour

Amok is a Malay term that means to engage furiously in battle. Clinically *amok* refer to an acute outburst of unrestrained violence associated with (indiscriminate) homicidal attacks, preceded by a period of brooding and ending with exhaustion and amnesia.[13] *Amok* homicides are distinct from other murders: the killer chooses an extremely destructive weapon, a crowded location, and insanely and indiscriminately kills a large number of people.[14]

There has been much speculation as to why *amok* behaviour tends to occur in Malay society. One explanation is its connection to the religious background of the people. Muslims are not permitted to commit suicide, which is considered a most heinous act in the Mohammedan religion. In the past, aggressive-homicidal behaviour influenced by infectious diseases has been considered, along with malaria, dengue, neurosyphilis, epilepsy, and so on, as biological in some cases. From a psychological point of view, an extraordinary sensitivity to hurt and the tendency to blame others for one's own difficulties are considered possible causes for the phenomenon. Loss of social standing, by way of insult, loss of employment, or financial loss, has been posited as a precipitating event for *amok*.

(ii) Amok behaviour in other areas

The outburst of aggressive (mass) homicidal behaviour is not necessarily confined to one cultural area, but can potentially be observed elsewhere, such as the territory of Papua and New Guinea, or many areas in South Asia, such as Thailand, the Philippines, Lao, and Indonesia.[5] It has challenged the previous view that *amok* occurs endemically within a particular society. It has been indicated that *amok* could happen in a fashion by communicability and through transmission from one population to another.[5] During the past decade in the United States, there have been increasing episodes of massive (and aimless) killing of people in neighbourhoods, workplaces, and of teachers/students in schools by deadly military weapons. These are American versions of *amok* attacks.

(b) Family suicide

When adults encounter severe difficulties (such as financial debt or a disgraceful event), there are many ways to deal with such problems. As one of the ways to cope with the difficult situation, Japanese parents, may decide to commit suicide together with their young children. This stress-coping method is based on the cultural

belief that it would be disgraceful to live after a shameful thing had happened, and that the shame would be relieved by ending one's life. This is coupled with the belief that the children, if left as orphans (after parents' committing husband–wife double suicide), would be mistreated by others. 'Blood is thicker than water' is the common saying reflecting the emphasis of blood-related family tie. Therefore, it would be better for them to die with their parents. This unique way of solving problems was often observed in Japan in the past. It is declining now but still observed some times.

(c) Cargo-cult syndromes (millenniary delusions)

Numerous social and behaviour scientists have noted that, historically, there have been occurrences of crisis cults in many different countries. The Taiping (Great Peace) Rebellion in China, Kikuyu maumau in Kenya, and the Ghost Dance of the Plains Indians of North America are some examples. Central to all these cultures are marked feelings of inferiority, conflict, and anxiety among the member-participants after being exposed to other, superior cultures and an attempt to renovate their self-images. Underlying these non-logical, magic-religious endeavours is a strong wish for resolution of their social, economic, and political problems and for a new and better way of life, such as that of the invading, superior cultures.

One kind of crisis cult is the cargo cult that has repeatedly arisen in Melanesia over the past century as a means of obtaining the manufactured articles possessed by European invaders.[15] Cargo is a neo-Melanesian or pidgin word that designates all of the manufactured goods, including canned foods and weapons, possessed by the Europeans, which are greatly desired by the indigenous people. Without knowing how the cargo was manufactured in the home countries of Europe, based on their own folk beliefs, the local people thought that it was given to the white people by their powerful ancestors through the performance of proper rituals. Accordingly, the local people tried to perform the white people's rituals, in the hopes that their ancestors would send them a lot of cargo and their lives would eventually be full of wealth. This behaviour might be individual, or it might involve a group of followers that gave up their normal lives to perform religious rituals, waiting for the arrival of the cargo, not only for several months, but for many years. They would become collectively deluded and led by a cult leader. As a culture-related specific syndrome, it may be understood that culture contributed to the stress that was encountered and also shaped the unique, pathological pattern of coping with it, a combination of pathogenic and pathoselective effects.

Culture-shaped variations of psychopathology (pathoplastic effect)

This category includes a group of disorders that manifests a clinical picture that is considerably different from the ordinal symptomatology of identified disorders described in current psychiatric classifications (of Euro-American origin). It is considered that the uniqueness of the symptomatology may be culturally attributed, i.e. due to cultural pathoplastic effects. Culture affects not only the content of symptoms, but, even more, the total clinical picture by the absence, addition, or variation of symptoms, resulting in considerable change in the manifestations of variations or subtypes of universally recognized psychiatric disorders.

(a) Anthropophobia (interpersonal relation phobia)

(i) Definition and nature of the disorder

Anthropophobia is the English translation for the Japanese term *taijin-kyofu-shio*. In Japanese, *taijin* means interpersonal, *kyofu* means phobic, and *shio* means syndrome or disorder. Therefore, *taijin-kyofu-shio* literally means the disorder with fear of interpersonal relation.[16] *Taijinkyofushio* is said to be prevalent among Japanese and is considered a culture-related psychiatric disorder. According to the clinical study, the onset of illness was as early mostly between 15 and 25 years, more prevalent among males than female. The cardinal symptoms manifested by the patients are: fear of one's bodily odours, fear of flushing, fear of showing odd attitudes towards others, fear of eye contact with others, concern about others' attitudes towards oneself, and fear of body dysmorphia.

(ii) Dynamic interpretation and culture formulation

The characteristic of *taijinkyofushio* is that the fear is induced in the presence of classmates, colleagues, or friends, those who are neither particularly close (such as family members) nor totally strange (such as people in the street). In other words, subjects are concerned with how to relate to people of intermediate familiarity. It is towards these people that a person must exercise delicate social etiquette. This is different from social phobia described in western societies, where patients have fear of socializing with strangers.

Culturally it has been explained that Japan is a situation-oriented society, very much concerned with how others see one's behaviour. It is considered that the act of staring at the person to whom one is talking is quite extraordinary and considered to be rude. Thus, there are cultural characteristics that cause Japanese to be hypersensitive about looking at and being looked at.

Taijinkyofushio is a psychological disorder of the adolescent. It is closely related to the problems associated with psychological development in the area of socialization. The Japanese child is raised in an atmosphere of indulgence and trust. However, when this protected child enters the wider world of junior high school, he or she faces multiple tasks: coping with conflict between biological needs and social restrictions, personal identity problems, and an increasing need for acceptance and love by others in social settings. This intensifies a feeling of unworthiness, making him or her more concerned about others' sensibilities and reactions.

(b) Brain fag syndrome

This syndrome was described originally as a very common minor psychiatric disorder occurring among the students of Southern Nigeria. Clinically, it is characterized by subjective complaints of intellectual impairment, (visual) sensory impairment, and somatic complaints, mostly of pain or a burning sensation in the head and neck. The student-patients often used the term brain fag to complain that they were no longer able to read, grasp what they were reading, or recall what they had just read, basically stressing their difficulty in mental function. Therefore, the term brain fag syndrome was suggested for this distinct clinical mental condition.[17]

In Nigeria, education was often a family affair, in which one of the brighter children was supported financially by family members; the educated member in turn was expected to be responsible for other family members when the need arose. Because of this family aspect of education, the student was burdened by the responsibility of maintaining the family's prestige. Thus, his or her academic

success or failure was associated with great stress. Therefore, it can be said that this syndrome is adjustment disorder with somatic feature and its symptomatology is shaped by culture.

(c) Arctic hysteria (pibloktoq)

Arctic (or polar) hysteria, also known by the local name *pibloktoq*, refers to a unique hysterical attack observed among the polar Eskimo people living in Arctic areas. The clinical condition is characterized by the sudden onset of loss or disturbance of consciousness. During the attack, as the patient may show various abnormal behaviours, such as tearing off his or her clothing, glossolalia, fleeing (nude or clothed), rolling in the snow, throwing anything handy around, performing mimetic acts, convulsion, or other bizarre behaviour. This emotional outburst occurs predominantly in women, but occasionally among men.[18]

No specific precipitating causes are noted. It has been speculated that the reaction is a manifestation of the basic Eskimo personality. Because the reaction is prevalent in winter, it is also thought that it may be related to increased threats of starvation or higher accident rates. Generally, it is suspected that the disorder is due to some basic, underlying anxiety, triggered by severe, culturally typical stresses: fear of certain impending situations, fear of loss, or fear of losing emotional support, including the sense of being on safe, solid, familiar ground.

(d) Malignant anxiety

A special, intensified form of anxiety disorder, termed as malignant anxiety, was reported to occur in Africa.[19] The condition was characterized by intense anxiety, extreme irritability, restlessness, and intense fear, and, therefore, named it malignant anxiety. It was referred to as frenzied anxiety as well. Often, the patient claimed that there was a change in his sense of self and reality, but there was no sign of personality deterioration or disintegration and no latent or overt psychotic symptoms. However, patients often suffered from intense feelings of anger that led to homicidal behaviour. The condition usually occurred in sporadic cases, but occasionally as an endemic. The disorder was thought to be situation related, associated with problems of adaptation to new and stressful life situations. Very often the patients were culturally marginal persons, who were in the process of renouncing their age-old cultures, but had failed to assimilate the new.

Culturally elaborated unique behaviour (pathoelaborating effect)

(a) Latah (startle-induced dissociative reaction)

(i) Defining the condition

Latah is a Malay word referring to the condition in which a person, after being startled by external stimuli, such as being tickled, suddenly experiences an altered consciousness and falls into a transient, dissociated state, exhibiting unusual behaviour (such as echolalia, echopraxia, or command automatism), including explosive verbal outbursts, usually of erotic words that are not ordinarily acceptable.[20] Beside Malaysia, the phenomenon has been observed in other places around the world and has been given various folk names: in Burma (where it is called *yaun*), Indonesia, Thailand (*bah-tsche*), the Philippines (*mali-mali*), indigenous tribes in Siberia, Russia (*myriachit*), and among the Ainu in Japan (who call it *imu*).

The *latah* reaction is found predominantly among women, although men may occasionally be involved. It was found that *latah* tends to run in families. In the past, it was primarily young women who were involved in *latah*. Most of the subjects found now are beyond middle age. Most cases are found in rural areas. Some develop *latah* reactions insidiously, without any precipitating events, whereas others symptoms occur after they endure psychologically stressful events. The loss of a significant person usually occurs shortly before the first experience of *latah* reaction. Once the reaction is experienced, it becomes habitual, and, thereafter, any sudden stimulation may provoke it. Hearing a sudden noise or being suddenly touched or poked by others may cause *latah*. Throwing a rope or other snake-like object in front of the person or simply shouting snake! will sufficiently startle the person to start a *latah* reaction.

The condition may last for several minutes or several hours if the person is continuously provoked. After the dissociative reaction is over, the subject usually claims amnesia and is puzzled about what had happened. Often, the subject is very apologetic and embarrassed for the (socially) inappropriate things he or she may have said (sexually coloured, dirty words) or done (such as touching men) during the attack.

(ii) Aetiological speculations

Anthropologists have tried to understand the *latah* condition from a cultural point of view. It has been pointed out that the *latah* reaction is remarkably congruent with the cultural themes emphasized, but in a paradoxical way. It is a peculiarly appropriate means of communicating marginality to others. The *latah* subjects are engaging in a performance, a role, and theatre, a culture-specific idiom expressing marginality while simultaneously reaffirming normative boundaries. The traditional polygamous Malay extended family structure is male dominated. Within this cultural system, *latah* is socially accepted as a female attention-seeking response, one of the few permissible overt, excitable, aggressive, and/or sexual demonstrations. In other words, *latah* is a culturally sanctioned emotional outlet for females. It is also a culturally elaborated unique behaviour.[21]

Cultural influence of prevalent occurrence of disorders (pathofacilitating effects)

This category includes several conditions that are commonly known as psychiatric disorders. There is nothing particular about them in terms of clinical manifestation (thus, they are not *specific* or *unique* disorders). However, their prevalence is influenced strongly enough by cultural factors that they may be viewed as heavily culture-related syndromes, rather than merely *ordinary* psychiatric disorders. Among this group of disorders, massive hysteria, group suicide, alcohol-related problems, or substance abuse are some of the examples.

Cultural interpretation of certain mental conditions (pathoreactive effect)

(a) Ataques de nervios (attack of nerves)

The folk name, *ataques de nervios*, literally meaning attack of nerves, refers to a stress-induced, culturally shaped unique emotional reaction with mixed anxiety-hysterical features.[22] This is an illness category used frequently by Hispanic people.

Initially observed among Puerto Rican army recruits, it was also labelled Puerto Rican syndrome. This condition typically occurs at funerals, in accidents, or in family conflicts, and calls forth family or other social supports. Commonly experienced symptoms include shaking, palpitations, a sense of heat rising to the head, and numbness, symptoms resembling a panic attack. The individual may shout, swear, and strike out at others, and finally fall to the ground, manifesting convulsion-like movements.

Based on clinical study, it has been reported that most of the patients (about 80 per cent) were female. From a clinical, diagnostic point of view, according to DSM-III-R criteria, the condition belongs to many subtypes of disorders, including panic disorder, recurrent major depression, generalized anxiety disorder, non-specific anxiety disorder, and others. Because the clinical picture is of a mixed, rather than a specific, nature, it may be interpreted simply as a folk label for an emotional reaction based on psychoreactive effect. It may be understood as an acute episode of social and psychological distress related to upsetting or frightening events in the family sphere. Focusing on symptoms alone misses what is most salient and meaningful about illness categories.

(b) Hwabyung (fire sickness)

Hwa-byung in the Korean language literally means fire (*hwa*) sickness (*byung*). Based on a traditional Chinese medical concept that is still prevalent in Korea, that an imbalance among the five elements within the body (metal, wood, water, fire, and earth) may cause physical disorders, laypeople in Korea use the folk term fire sickness to describe ill conditions. This is a folk idiom of distress characterized by a wide range of somatic and emotional symptoms. About three-fourths of the patients that complained of *hwabyung* were women, who linked their conditions to anger provoked by domestic problems, such as their husbands' extramarital affairs and strained in-law relationships. Culturally, it has been explained that male chauvinism has always been dominant in Korean society, and women tend to suffer from their vulnerable status. When a housewife is mistreated by her husband or is having troubles with her in-laws, she has to suppress her emotional reactions so that there will be no disturbance in the stability of the family. As a woman, she is taught to accept defeat, bear frustration, and suppress her hatred. As a result, accumulated resentment (*hahn* in Korean) becomes a major issue for some women. This is the core dynamic for understanding *hwabyung*.[23]

Cultural factors may indirectly contribute to the occurrence of particular psychological problems that are encountered by Korean women, but through pathoreactive effect, emotional reaction was labelled by laymen as an indication of fire sickness (*hwabyung*).

(c) Susto (soul loss)

Susto is a Spanish word that literally means fright. The term is widely used by people in Latin America to refer to the condition of loss of soul.[24] It is based on the folk belief that every individual possesses a soul, but, through certain experiences, such as being frightened or startled, a person's soul may depart from the body. As a result, the soul-lost person will manifest certain morbid mental conditions and illness behaviour. The remedy for such a condition is to recapture the soul through certain rituals. The concept of loss of soul as a cause for sickness is prevalent around the world, and that terms similar, or equivalent, to *susto* are found widely distributed across many different cultural groups, such as *el miedo* (fright) in Bolivia, *lanti* in the Philippines, or *mogo laya* in Papua New Guinea.

It should be pointed out that, although the cause is attributed to spiritual-psychological reasons relating mostly to a frightening experience or misfortune, from a clinical point of view, the manifested syndrome is quite heterogeneous, without a commonly shared syndrome. The victim may manifest loss of appetite, sleep disturbance, reduced strength, absentmindedness, headache, dizziness, or other somatic symptoms, as well as emotional symptoms of depression, anxiety, or irritability. Therefore, strictly speaking, it is not a culture-related specific syndrome derived from psychogenic or psychoplastic effects. It is culture-related *only* in the sense that the morbid condition is interpreted after the fact according to folk concepts of aetiology and certain ways of regaining the lost soul, such as rituals, are offered. Therefore, the role of culture is interpretation of and reaction to the illness.

Final comments

Diagnosis and classification issues

Associated with the increased awareness of the impact of culture on psychiatric classifications, there is controversy regarding how to deal with culture-related specific syndromes from a diagnostic point of view.[12] Some clinicians feel strongly that various known culture-bound syndromes (such as *koro* or *hwabyung*) should be officially recognized and included in the present classification system.

However, it needs to be pointed out that the present DSM classification system is based on the descriptive approach, categorizing psychiatric disorders by certain sets of manifested symptomatology. If we try to fit culture-related specific syndromes into the categories of the existing classification system or try to create new categories of disorders, they will be classified as NOS (not otherwise specified) or, at best, as variations of presently recognized disorders. Most importantly, by squeezing the culture-related specific syndromes into the descriptive-oriented classification system, we will lose the unique meaning of the syndromes from a cultural perspective.[25]

Culture-related specific syndromes in western societies

Another point that must be made is that, by definition, culture-related specific syndromes should be able to be discovered everywhere, as every society, no matter East or West, has its own culture. However, the trend has been to consider that most culture-related specific syndromes (such as *koro*, *amok*, or *dhat* syndromes) occur in non-western societies. This is because they were considered peculiar phenomena observed in areas previously colonized by western people and they simply did not fit the classification system developed for Euro-American populations. This trend is now changing. There is an increased interest in recognizing syndromes in our own western cultures that are heavily culture-related.[26] Several psychiatric disorders have been suggested by various scholars for consideration as western culture-related syndromes. These include: anorexia nervosa, obesity, drug-induced dissociated states, and multiple personality, disorders that are seldom observed or concerned in non-western societies. Because these conditions are already recognized in the existing western nosological system, they are, in a sense, not viewed as specific syndromes. However,

they can be viewed as culture-related psychiatric conditions that are influenced by the pathoelaborative, pathofacilitative, or pathoplastic effects of western culture.

Clinical implications for general psychiatry

Even though the encounter of culture-related specific psychiatric disorder in our daily psychiatric practice is relatively rare, the purpose of examining such specific syndromes has its significant purpose and implications. Through such unique examples, it helps us to appreciate the cultural attribution to the stress formation, reaction pattern, symptom manifestation, occurrence of frequency of disorders, and reaction to the disorders. It also concerns how to work on therapy for the disorder by complying patient's cultural background.

Further information

Tseng, W.S. (2001). D: 13. Culture-related specific syndromes. In *Handbook of cultural psychiatry* (ed. W.S. Tseng), pp. 211–63. Academic Press, San Diego.

References

1. Yap, P.M. (1967). Classification of the culture-bound reactive syndromes. *The Australian and New Zealand Journal of Psychiatry*, **1**, 172–9.
2. Jilek, W.G. and Jilek-Aall, L. (2001). Culture-specific mental disorders. In *Contemporary psychiatry. Psychiatry in special situations*, Vol. 2 (eds. F. Henn, N. Sartorius, H. Helmchen, *et al.*), pp. 219–45. Springer, Berlin.
3. Tseng, W.S. (2001). *Handbook of cultural psychiatry*. Academic Press, San Diego.
4. Tseng, W.S. (2006). From peculiar psychiatric disorders through culture-bound syndromes to culture-related specific syndromes. *Transcultural Psychiatry*, **43**, 554–76.
5. Westermeyer, J. (1973). On the epidemicy of *amok* violence. *Archives of General Psychiatry*, **28**, 873–6.
6. Prince, R. and Tcheng-Laroche, F. (1987). Culture-bound syndromes and international disease classifications. *Culture, Medicine and Psychiatry*, **11**, 3–19.
7. Chowdhury, A.N. (1994). Koro in females: an analysis of 48 cases. *Transcultural Psychiatric Research Review*, **31**, 369–80.
8. Tseng, W.S., Mo, G.M., Hsu, J., *et al.* (1988). A sociocultural study of koro epidemics in Guandong, China. *The American Journal of Psychiatry*, **145**, 1538–43.
9. Malhotra, H.K. and Wig, N.N. (1975). Dhat syndrome: a culture-bound sex neurosis of the orient. *Archives of Sexual Behavior*, **4**, 519–28.
10. Neki, J.S. (1973). Psychiatry in south-east Asia. *The British Journal of Psychiatry*, **123**, 256–69.
11. Bhatia, M.S. and Malik, S.C. (1991). Dhat syndrome—a useful diagnostic entity in Indian culture. *The British Journal of Psychiatry*, **159**, 691–5.
12. Hughes, C.C. (1996). The culture-bound syndromes and psychiatric diagnosis. In *Culture and psychiatric diagnosis*: *a DSM-IV perspective* (eds. J.E. Mezzich, A. Kleinman, H. Fabrega, Jr, *et al.*), pp. 289–307. American Psychiatric Press, Washington, DC.
13. Yap, P.M. (1951). Mental diseases peculiar to certain cultures: a survey of comparative psychiatry. *The Journal of Mental Science*, **97**, 313–27.
14. Westermeyer, J. (1972). A comparison of *amok* and other homicide in Laos. *The American Journal of Psychiatry*, **129**, 703–9.
15. Burton-Bradley, B.G. (1975). Cargo cult. In *Stone age crisis: a psychiatric appraisal* (ed. B.G. Burton-Bradley), pp. 10–31. Vanderbilt University Press, Nashville, TN.
16. Kitanishi, K. and Mori, A. (1995). Morita therapy: 1919 to 1995. *Psychiatry and Clinical Neurosciences*, **13**, 31–7.
17. Prince, R. (1960). The brain fag syndrome in Nigerian students. *The Journal of Mental Science*, **104**, 559–70.
18. Foulks, E.F. (1972). The Arctic hysterias of the North Alaskan Eskimo. In *Anthropological studies* (No. 10) (ed. D.H. Maybury Lewis). American Anthropological Association, Washington, DC.
19. Lambo, T.A. (1962). Malignant anxiety: a syndrome associated with criminal conduct in Africans. *The Journal of Mental Science*, **108**, 256–64.
20. Simons, R.C. (1980). The resolution of the latah paradox. *The Journal of Nervous and Mental Disease*, **168**, 195–206.
21. Simons, R.C. (1996). *Boo!—culture, experience, and the startle reflex*. Oxford University Press, New York.
22. Guarnaccia, P.J. (1993). Ataques de nervios in Puerto Rico: culture-bound syndrome or popular illness? *Medical Anthropology*, **15**, 157–70.
23. Min, S.W.K. (2006). Hwabyung: a culture-related chronic anger syndrome of Korea. The First World Congress of Cultural Psychiatry Proceeding. (S-III-23).
24. Rubel, A.J. (1964). The epidemiology of a folk illness: *Susto* in Hispanic America. *Ethology*, **3**, 268–83.
25. Pfeiffer, W.M. (1982). Culture-bound syndromes. In *Culture and psychopathology* (ed. I. Al-Issa). University Park Press, Baltimore.
26. Littlewood, R. and Lipsedge, M. (1986). The culture-bound syndromes of the dominant culture: culture, psychopathology and biomedicine. In *Transcultural psychiatry* (ed. J.L. Cox), pp. 253–73. Croom Helm, London.

This page intentionally left blank

New Oxford Textbook of

Psychiatry

This page intentionally left blank

VOLUME 2

New Oxford Textbook of Psychiatry

SECOND EDITION

Edited by

Michael G. Gelder
Emeritus Professor of Psychiatry,
Warneford Hospital, University of Oxford,
Oxford, UK

Nancy C. Andreasen
Director, Mental Health Clinical Research Centre,
University of Iowa Hospital and Clinic,
Iowa City, USA

Juan J. López-Ibor Jr.
Professor of Psychiatry,
Complutense University, Madrid, Spain

and

John R. Geddes
Professor of Epidemiological Psychiatry
University of Oxford, Warneford Hospital,
Oxford, UK

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford ox2 6dp

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First edition published 2000
Reprinted 2003

This edition published 2009

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloguing in Publication Data

Data available

Typeset in Cepha Imaging Pvt. Ltd., Bangalore, India
Printed in Italy on acid-free paper by Rotolito Lombarda SpA

ISBN 978-0-19-920669-8
ISBN 978-0-19-920669-8 (set)
ISBN 978-0-19-955992-3 (Volume 1)
ISBN 978-0-19-955993-0 (Volume 2)

10 9 8 7 6 5 4 3 2 1

Preface to the second edition

This new edition, like the first, aims to present a comprehensive account of clinical psychiatry with reference to its scientific basis and to the ill person's perspective. As in the first edition, the authors are drawn from many countries, including the UK, the USA, 12 countries in continental Europe, and Australasia. The favourable reception of the first edition has led us to invite many of the original authors to revise their chapters for this second edition but 50 chapters are the work of new authors, many concerned with subjects that appeared in the first edition, while others are completely new. The forensic psychiatry section has the most new chapters, followed by the section on psychology as a scientific basis of psychiatry.

The overall plan of the book resembles that of the first edition (see preface to the 1st edition, reprinted on pages vii and viii). One important feature is that information about treatment appears in more than one place. The commonly used physical and psychological treatments are described in Section 6. Their use in the treatment of any particular disorder is considered in the chapter concerned with that disorder and the account is in two parts. The first part is a review of evidence about the effects of each of the treatments when used for that disorder. The second part, called Management, combines evidence from clinical trials with accumulated clinical experience to produce practical advice about the day to day care of people with the disorder.

Although much information can now be obtained from internet searches, textbooks are still needed to provide the comprehensive account of established knowledge into which new information can be fitted and against which recent findings can be evaluated. As well as seeking to provide an authoritative account of essential knowledge, each chapter in the new edition includes a brief list of sources of further information, including where appropriate, regularly updated web sites.

An essential component of good practice is the need to be aware of patients' perspectives, to respect their wishes, and to work with them, and often their families, as partners. The book opens with an important chapter on the experience of being a patient, and there are chapters on stigma, ethics, and the developing topic of values-based practice.

We are grateful to the following who advised us about parts of the book; Professor John Bancroft (Psychosexual Disorders), Professor Tom Burns (Social and Community Psychiatry), Professor William Fraser (Intellectual Disability), Professor Keith Hawton (Suicide and Deliberate Self Harm), Professor Susan Iversen (Psychology), Professor Robin Jacoby (Old Age Psychiatry), Professor Paul Mullen (Forensic Psychiatry), Sir Michael Rutter (Child and Adolescent Psychiatry), and Professor Gregory Stores (Sleep Disorders).

The editors

This page intentionally left blank

Preface to the first edition

Three themes can be discerned in contemporary psychiatry: the growing unity of the subject, the pace of scientific advance, and the growth of practice in the community. We have sought to reflect these themes in the *New Oxford Textbook of Psychiatry* and to present the state of psychiatry at the start of the new millennium. The book is written for psychiatrists engaged in continuous education and recertification; the previous, shorter, *Oxford Textbook of Psychiatry* remains available for psychiatrists in training. The book is intended to be suitable also as a work of reference for psychiatrists of all levels of experience, and for other professionals whose work involves them in the problems of psychiatry.

The growing unity of psychiatry

The growing unity in psychiatry is evident in several ways. Biological and psychosocial approaches have been largely reconciled with a general recognition that genetic and environmental factors interact, and that psychological processes are based in and can influence neurobiological mechanisms. At the same time, the common ground between the different psychodynamic theories has been recognized, and is widely accepted as more valuable than the differences between them.

The practice of psychiatry is increasingly similar in different countries, with the remaining variations related more to differences between national systems of health care and the resources available to clinicians, than to differences in the aims of the psychiatrists working in these countries. This unity of approach is reflected in this book whose authors practise in many different countries and yet present a common approach. In this respect this textbook differs importantly from others which present the views of authors drawn predominantly from a single country or region.

Greater agreement about diagnosis and nosology has led to a better understanding of how different treatment approaches are effective in different disorders. The relative specificity of psychopharmacological treatments is being matched increasingly by the specificity of some of the recently developed psychological treatments, so that psychological treatment should no longer be applied without reference to diagnosis, as was sometimes done in the past.

The pace of scientific advance

Advances in genetics and in the neurosciences have already increased knowledge of the basic mechanisms of the brain and are beginning to uncover the neurobiological mechanisms involved in psychiatric disorder. Striking progress has been achieved in the understanding of Alzheimer's disease, for example, and there are indications that similar progress will follow in uncovering the causes of mood disorder, schizophrenia, and autism. Knowledge of genetics and the neurosciences is so extensive and the pace of change is so rapid that it is difficult to present a complete account within the limited space available in a textbook of clinical psychiatry. We have selected aspects of these sciences that seem, to us and the authors, to have contributed significantly to psychiatry or to be likely to do so before long.

Psychological and social sciences and epidemiology are essential methods of investigation in psychiatry. Although the pace of advance in these sciences may not be as great as in the neurosciences, the findings generally have a more direct relation to clinical phenomena. Moreover, the mechanisms by which psychological and social factors interact with genetic, biochemical, and structural ones will continue to be important however great the progress in these other sciences. Among the advances in the psychological and social sciences that are relevant to clinical phenomena, we have included accounts of memory, psychological development, research on life events, and the effects of culture. Epidemiological studies continue to be crucial for defining psychiatric disorders, following their course, and identifying their causes.

Psychiatry in the community

In most countries, psychiatry is now practised in the community rather than in institutions, and where this change has yet be completed, it is generally recognized that it should take place. The change has done much more than transfer the locus of care; it has converted patients from passive recipients of care to active participants with individual needs and preferences. Psychiatrists are now involved in the planning, provision, and evaluation of services for whole communities, which may include members of ethnic minorities, homeless people, and refugees. Responsibility for a community has underlined the importance of the prevention as well as the treatment of mental disorder and of the role of agencies other than health services in both. Care in the community has also drawn attention to the many people with psychiatric disorder who are treated in primary care, and has led to new ways of working between psychiatrists and physicians. At the same time, psychiatrists have

worked more in general hospitals, helping patients with both medical and psychiatric problems. Others have provided care for offenders.

The organization of the book

In most ways, the organization of this book is along conventional lines. However, some matters require explanation.

Part 1 contains a variety of diverse topics brought together under the general heading of the subject matter and approach to psychiatry. Phenomenology, assessment, classification, and ethical problems are included, together with the role of the psychiatrist as educator and as manager. Public health aspects of psychiatry are considered together with public attitudes to psychiatry and to psychiatric patients. Part 1 ends with a chapter on the links between science and practice. It begins with a topic that is central to good practice—the understanding of the experience of becoming a psychiatric patient.

Part 2 is concerned with the scientific foundations of psychiatry grouped under the headings neurosciences, genetics, psychological sciences, social sciences, and epidemiology. The chapters contain general information about these sciences; findings specific to a particular disorder are described in the chapter on that disorder. Brain imaging techniques are discussed here because they link basic sciences with clinical research. As explained above, the chapters are selective and, in some, readers who wish to study the subjects in greater detail will find suggestions for further reading.

Part 3 is concerned with dynamic approaches to psychiatry. The principal schools of thought are presented as alternative ways of understanding the influence of life experience on personality and on responses to stressful events and to illness. Some reference is made to dynamic psychotherapy in these accounts, but the main account of these treatments is in Part 6. This arrangement separates the chapters on the practice of dynamic psychotherapy from those on psychodynamic theory, but we consider that this disadvantage is outweighed by the benefit of considering together the commonly used forms of psychotherapy.

Part 4 is long, with chapters on the clinical syndromes of adult psychiatry, with the exception of somatoform disorders which appear in Part 5, Psychiatry and Medicine. This latter contains more than a traditional account of psychosomatic medicine. It also includes a review of psychiatric disorders that may cause medical symptoms unexplained by physical pathology, the medical, surgical, gynaecological, and obstetric conditions most often associated with psychiatric disorder, health psychology, and the treatment of psychiatric disorder in medically ill patients.

Information about treatment appears in more than one part of the book. Part 6 contains descriptions of the physical and psychological treatments in common use in psychiatry. Dynamic psychotherapy and psychoanalysis are described alongside counselling and cognitive behavioural techniques. This part of the book contains general descriptions of the treatments; their use for a particular disorder is considered in the chapter on that disorder. In the latter, the account is generally in two parts: a review of evidence about the efficacy of the treatment, followed by advice on management in which available evidence is supplemented, where necessary, with clinical experience. Treatment methods designed specially for children and adolescents, for people with mental retardation (learning disability), and for patients within the forensic services are considered in Parts 9, 10, and 11 respectively.

Social psychiatry and service provision are described in Part 7. Public policy issues, as well as the planning, delivery, and evaluation of services, are discussed here. Psychiatry in primary care is an important topic in this part of the book. There are chapters on the special problems of members of ethnic minorities, homeless people, and refugees, and the effects of culture on the provision and uptake of services.

Child and adolescent psychiatry, old age psychiatry, and mental retardation are described in Parts 8, 9, and 10. These accounts are less detailed than might be found in textbooks intended for specialists working exclusively in the relevant subspecialty. Rather, they are written for readers experienced in another branch of psychiatry who wish to improve their knowledge of the special subject. We are aware of the controversy surrounding our choice of the title of Part 10. We have selected the term 'mental retardation' because it is used in both ICD-10 and DSM-IV. In some countries this term has been replaced by another that is thought to be less stigmatizing and more acceptable to patients and families. For example, in the United Kingdom the preferred term is 'learning disability'. While we sympathize with the aims of those who adopt this and other alternative terms, the book is intended for an international readership and it seems best to use the term chosen by the World Health Organization as most generally understood. Thus the term mental retardation is used unless there is a special reason to use another.

In Part 11, Forensic Psychiatry, it has been especially difficult to present a general account of the subject that is not tied to practice in a single country. This is because systems of law differ between countries and the practice of forensic psychiatry has to conform with the local legal system. Although many of the examples in this part of the book may at first seem restricted in their relevance because they are described in the context of English law, we hope that readers will be able to transfer the principles described in these chapters to the legal tradition in which they work.

Finally, readers should note that the history of psychiatry is presented in more than one part of the book. The history of psychiatry as a medical specialty is described in Part 1. The history of ideas about the various psychiatric disorders appears, where relevant, in the chapters on these disorders, where they can be considered in relation to present-day concepts. The history of ideas about aetiology is considered in Part 2, which covers the scientific basis of psychiatric aetiology, while the historical development of dynamic psychiatry is described in Part 3.

Michael Gelder
Juan López-Ibor
Nancy Andreasen

Acknowledgements from the first edition

We are grateful to the many colleagues who have advised us about certain parts of the book.

The following helped us to plan specialized parts of the book: Dr Jeremy Holmes (Section 3, Psychodynamic Contributions to Psychiatry); Professor Richard Mayou (Section 5, Psychiatry and Medicine); Professor Robin Jacoby (Section 8, Psychiatry of Old Age); Sir Michael Rutter (Section 9, Child and Adolescent Psychiatry); Professor William Fraser (Section 10, Intellectual Disablity); Professor Robert Bluglass (Section 11, Forensic Psychiatry).

The following helped us to plan certain sections within Section 4, General Psychiatry: Professor Alwyn Lishman (delirium, dementia, amnestic syndrome, and other cognitive disorders); Professor Griffith Edwards (alcohol use disorders); Dr Philip Robson (other substance use disorders); Professor Guy Goodwin (mood disorders); Professor John Bancroft (sexuality, gender identity, and their disorders); Professor Gregory Stores (sleep–wake disorders); Professor Keith Hawton (suicide and attempted suicide). In Section 6, Professor Philip Cowen advised about somatic treatments, Dr Jeremy Holmes about psychodynamic treatments, and Professor David Clark about cognitive behavioural therapy. Dr Max Marshall provided helpful advice about forensic issues for Section 7. We also thank the many other colleagues whose helpful suggestions about specific problems aided the planning of the book.

Finally, we record our special gratitude to the authors and to the staff of Oxford University Press.

This page intentionally left blank

Contents
Volume 1

Contents
Volume 2

This page intentionally left blank

Contributors list

S.E. Abbey Associate Professor of Psychiatry, University of Toronto, Toronto, Canada
Chapter 5.3.6

Henry David Abraham Distinguished Life Fellow, American Psychiatric Association
Chapter 4.2.3.3

Clive E. Adams Cochrane Schizophrenia Group, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapter 6.1.1

Hagop S. Akiskal Professor of Psychiatry and Director of the International Mood Center, University of California at San Diego, California, USA
Chapter 4.5.9

Nancy C. Andreasen Dept of Psychiatry, University of Iowa Hospitals & Clinics, Iowa City, USA
Chapter 4.3.1

Adrian Angold Associate Professor of Child and Adolescent Psychiatry, Duke University Medical Center, Durham, North Carolina, USA
Chapter 9.1.2

Jules Angst Emeritus Professor of Psychiatry, Zurich University
Chapter 4.5.7

Ella Arensman Director of Research, National Suicide Research Foundation, Ireland
Chapter 4.15.2

J.K. Aronson Reader in Clinical Pharmacology, University Department of Primary Health Care, Headington, Oxford
Chapter 6.2.1

José-Luis Ayuso-Mateos Chairman, Department of Psychiatry, Universidad Autónoma de Madrid, Hospital Universitario de la Princesa, Spain
Chapter 5.3.5

Susan Bailey Consultant Child and Adolescent Forensic Psychiatrist, Salford NHS Trust and Maudsley NHS Trust; Senior Research Fellow, University of Manchester, UK
Chapter 11.7

Robert Baldwin Consultant, Old Age Psychiatrist, and Honorary Senior Lecturer, Manchester Royal Infirmary, UK
Chapter 8.5.4

James C. Ballenger Retired Professor and Chairman, Department of Psychiatry and Behavioral Sciences and Director, Institute of Psychiatry, Medical University of South Carolina
Chapter 4.7.3

John Bancroft, The Kinsey Institute for Research in Sex, Gender, & Reproduction and Department of Psychiatry, University of Oxford
Chapter 4.11.2

David H. Barlow Center for Anxiety and Related Disorders at Boston University, Massachusetts, USA
Chapter 4.7.1

Delwyn J. Bartlett Woolcock Institute of Medical Research, Sydney, Australia
Chapter 4.14.2

Christopher Bass Consultant in Liaison Psychiatry, John Radcliffe Hospital, Oxford, UK
Chapters 5.2.4 and 5.2.9

Antony W. Bateman Halliwick Psychotherapy Dept, St Ann's Hospital, London, UK
Chapter 4.12.6

Per Bech Professor of Psychiatry and Head of Psychiatric Research Unit, WHO Collaborating Centre, Frederiksborg General Hospital, Hillerød, Denmark
Chapter 4.5.2

Sidney Benjamin Senior Lecturer, University of Manchester, UK
Chapter 5.2.6

Thomas P. Berney Consultant Developmental Psychiatrist Honorary Research Associate, University of Newcastle upon Tyne
Chapter 10.6

Jose M. Bertolote Chief, Mental Disorders Control Unit, World Health Organization, Geneva; Associate Professor, Department of Psychogeriatrics, University of Lausanne, Switzerland
Chapter 7.4

Zubin Bhagwagar CT Mental Health Center, Yale University, New Haven CT, USA
Chapter 6.2.3

Mary Beth Connolly Gibbons Assistant Professor of Psychology in Psychiatry Department of Psychiatry, University of Pennsylvania, Pennsylvania, USA
Chapter 6.1.2

Dinesh Bhugra Professor of Mental Health and Cultural Diversity, King's College London, Institute of Psychiatry, London, UK
Chapter 7.10.3

Michel Billiard Professor of Neurology, School of Medicine, Guide Chauliac Hospital, Montpellier, France
Chapter 4.14.3

Max Birchwood Director, Early Intervention Service, Northern Birmingham Mental Health Trust, and University of Birmingham, UK
Chapter 6.3.2.4

Boris Birmaher UPMC Western Psychiatric Institute, Pittsburgh, USA
Chapter 9.2.7

Stella Bitran, Center for Anxiety and Related Disorders, Boston University, Beacon, MA
Chapter 4.7.1

Alan H. Bittles Centre for Comparative Genomics, Murdoch University, Perth, Australia
Chapter 8.1

Dora Black Honorary Consultant, Child and Adolescent Psychiatry, Traumatic Stress Clinic, London; Honorary Lecturer, University of London, UK
Chapters 9.3.2 and 9.3.7

Michelle A. Blackmore, Doctoral Student of Clinical Psychology Adult Anxiety Clinic at Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Carlos Blanco New York State Psychiatric Institute, New York,
Chapter 6.3.3

Sidney Bloch Professor of Psychiatry, University of Melbourne; Senior Psychiatrist, St Vincent's Hospital, Melbourne, Australia
Chapters 1.5 and 6.3.8

William V. Bobo Assistant Professor of Psychiatry, Vanderbilt University School of Medicine Nashville, Tennessee (USA)
Chapter 6.2.5

Bernhard Bogerts Department of Psychiatry, University of Magdeburg, Germany
Chapter 2.3.5

Jeff Bostic School of Psychiatry, Harvard Medical School, Cambridge MA, USA
Chapter 9.1.3

Nick Bouras Professor, Institute of Psychiatry - King's College London MHiLD - York Clinic, London, UK
Chapter 10.9

Christopher R. Bowie Department of Psychiatry, Mount Sinai School of Medicine, New York, USA
Chapter 4.3.3

Colin Bradbury Department of Psychological Medicine, Institute of Psychiatry, De Crespigny Park, London, UK
Chapter 4.2.4

Brian P. Brennan Instructor in Psychiatry, Harvard Medical School and Associate Director for Translational Neuroscience Research, Biological Psychiatry Laboratory, McLean Hospital, Belmont, MA
Chapter 6.2.6

David Brent Dept of Psychiatry, University of Pittsburgh Medical School, Pittsburgh PA, USA
Chapter 9.2.7

Chris R. Brewin Research Dept of Clinical, Educational & Health Psychology, University College London, UK
Chapter 4.6.3

Elaine Brohan Institute of Psychiatry, David Goldberg Centre, De Crespigny Park, London, UK
Chapter 1.2

Ian Brockington Professor of Psychiatry, University of Birmingham, UK
Chapter 5.4

George W. Brown Professor of Sociology, Academic Department of Psychiatry, St Thomas's Hospital, London, UK
Chapter 2.6.1

Arne Brun Professor of Neuropathology. Department of Pathology, Lund University Hospital, Lund, Sweden
Chapter 4.1.3

Richard A. Bryant School of Psychology, University of New South Wales, Sydney NSW, Australia
Chapter 4.6.1

E.T. Bullmore Institute of Psychiatry, King's College London, UK
Chapters 2.3.7 and 2.3.8

Tom Burns Professor of social psychiatry, Dept of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
Chapter 7.5

José Luis Carrasco Professor of Psychiatry, Hospital Fundacion Jimenez Diaz, Universidad Autonoma, Madrid, Spain
Chapter 4.12.3

D.J. Castle University of Western Australia, Fremantle, Australia
Chapter 4.3.6.1

Jonathan Chick Consultant Psychiatrist, NHS Lothian, and Senior Lecturer, Department of Psychiatry, University of Edinburgh
Chapter 4.2.2.4

Daniel Chisholm Department of Health System Financing, Health Systems and Services, World Health Organization, Geneva, Switzerland
Chapter 7.7

David Christmas Dept of Psychiatry, University of Dundee, Dundee, UK
Chapter 6.2.10.4

David M. Clarke Consultant Psychiatrist, Lea Castle Centre, Kidderminster DY10 3PP
Chapters 6.3.3.1 and 10.4

L. Clark Dept of Experimental Psychology, University of Cambridge, Cambridge, UK
Chapter 2.5.5

Judy Clegg Lecturer, Speech and language therapist, HPC, RCSLT Department of Human Communication Sciences University of Sheffield, UK
Chapter 9.2.11

C. Robert Cloninger Dept of Psychiatry, Washington University School of Medicine, St Louis MO, USA
Chapter 1.8.2

John Collinge Head of the Department of Neurodegenerative Disease at the Institute of Neurology, University College London and the Director of the UK Medical Research Council's Prion Unit, London, UK
Chapter 4.1.4

Henry O'Connell Consultant Psychiatrist, Co. Tipperary, Ireland
Chapter 8.5.2

Melissa A. Culhane Harvard Program in Refugee Trauma, Department of Psychiatry, Massachusetts General Hospital, Cambridge, USA
Chapter 7.10.1

John E. Cooper Emeritus Professor of Psychiatry, University of Nottingham, UK
Chapter 1.8.1

Sally-Ann Cooper Professor of Learning Disabilities, Division of Community Based Sciences, Faculty of Medicine, University of Glasgow
Chapter 10.2

Zafra Cooper Principal Research Psychologist, Oxford University Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapters 4.10.2 and 6.3.2.2

E. Jane Costello Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Brightleaf Square, Durham NC
Chapter 9.1.2

Elizabeth Coulthard Institute of Neurology, University College London, UK
Chapter 2.5.2

F.-R. Cousin Psychiatrist, Centre Hospitalier Saint-Anne, Paris, France
Chapter 4.3.10

Philip J. Cowen Professor of Psychopharmacology, Department of Psychiatry, University of Oxford
Chapter 6.2.10.2

Tom K.J. Craig Professor of Social Psychiatry, King's College London, Institute of Psychiatry, London UK
Chapters 7.10.2 and 7.10.3

Paul Crits-Christoph Professor of Psychology in Psychiatry Director, Center for Psychotherapy Research Department of Psychiatry, University of Pennsylvania. Pennsylvania, USA
Chapter 6.1.2

Michael Crowe Consultant Psychiatrist, South London and Maudsley NHS Trust; Honorary Senior Lecturer, Institute of Psychiatry, King's College London, UK
Chapter 6.3.7

D.G. Cunningham Owens Reader in Psychiatry, Department of Psychiatry, University of Edinburgh, UK
Chapter 4.3.8

Dianne Currier Division of Molectular Imaging & Neuropathology, Department of Psychiatry, Columbia University
Chapter 4.15.3

Robert Dantzer Integrative Immunology and Behavior Program, University of Illinois at Urbana-Champaign, Edward R. Madigan Laboratory, West Gregory Drive, Urbana, IL, USA
Chapter 2.3.10

Rajan Darjee Division of Psychiatry, University of Edinburgh, Edinburgh, UK
Chapter 11.3.1

Anthony S. David Professor of Cognitive Neuropsychiatry, Institute of Psychiatry, King's College London, UK
Chapters 4.3.4 and 4.9

Kate Davidson Senior Research Psychologist, Department of Psychological Medicine, University of Glasgow, UK
Chapter 4.12.7

Martin Davies Dept of Experimental Psychology, University of Oxford, Oxford, UK
Chapter 2.1

Giovanni de Girolamo Health Care Research Agency, Emilia-Romagna Region, Bologna, Italy
Chapters 4.12.2 and 4.12.4

Shoumitro Deb Clinical Professor of Neuropsychiatry & Intellectual Disability, Division of Neuroscience, University of Birmingham, UK
Chapter 10.4

R.J. Dolan Institute of Neurology, University College London, UK
Chapter 2.5.4

Anton Došen Emeritus Professor of Psychiatric Aspects of Intellectual Disability at the Radboud University, Nijmegen, The Netherlands
Chapter 10.5.2

D. Colin Drummond Professor of Addiction Psychiatry, Section of Alcohol Research, National Addiction Centre, Division of Psychological Medicine and Psychiatry, Institute of Psychiatry, King's College London
Chapter 4.2.2.5

Emma Dunn School of Medicine, Cardiff University, Cardiff, UK
Chapter 11.17

Graham Dunn Professor of Biomedical Statistics, Health Methodology Research Group, School of Community Based Medicine, University of Manchester
Chapter 2.2

Julie Dunsmore Honorary Clinical Associate, SciMHA Unit, University of Western Sydney, Australia
Chapter 4.6.5

Anke Ehlers Department of Psychiatry, University of Oxford, UK
Chapters 4.6.1 and 4.6.2

Timo Erkinjuntti Professor of Neurology, Head of the University Department of Neurological Sciences, University of Helsinki and Head Physician, Department of Neurology and Memory Research Unit, Helsinki University Central Hospital, Finland
Chapter 4.1.8

Brigette A. Erwin Adult Anxiety Clinic of Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Colin A. Espie Professor of Clinical Psychology and Head of Department of Psychological Medicine, University of Glasgow, UK
Chapter 4.14.2

Jonathan J. Evans Section of Psychological Medicine, University of Glasgow, Glasgow, UK
Chapter 4.1.14

Christopher G. Fairburn Wellcome Principal Research Fellow and Professor of Psychiatry, University of Oxford, UK
Chapters 4.10.2 and 6.3.2.2

Peter Falkai Professor of Medical Psychology, Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany
Chapter 2.3.5

Stephen V. Faraone Director, Medical Genetics Research, Professor of Psychiatry and of Neuroscience & Physiology, Director, Child and Adolescent Psychiatry Research, SUNY Upstate Medical University, New York
Chapter 4.3.9

Michael Farrell Senior Lecturer and Consultant Psychiatrist, National Addiction Centre, South London and Maudsley NHS Trust, London, UK
Chapter 4.2.3.4

David P. Farrington Professor of Psychological Criminology, University of Cambridge, UK
Chapter 11.2

J. Paul Fedoroff Director, Sexual Behaviors Clinic Royal Ottawa Mental Health Centre and Director of Forensic Research University of Ottawa Institute of Mental Health Research
Chapter 4.11.3

Howard H. Feldman Professor and Head, Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada
Chapter 8.5.1.1

Melanie J.V. Fennell Consultant Clinical Psychologist; Director, Oxford Diploma in Cognitive Therapy, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapter 6.3.2.3

Max Fink Emeritus Professor of Psychiatry and Neurology, State University of New York at Stony Brook; Professor of Psychiatry, Albert Einstein College of Medicine; Attending Psychiatrist, Long Island Jewish Medical Center, New York, USA
Chapter 6.2.10.1

Michael B. First Columbia University, New York, USA
Chapter 1.9

Per Fink Director, Research Unit for Functional Disorders, Aarhus University Hospital, Risskov, Denmark
Chapter 5.2.3

Philip A. Fisher Research Scientist, Oregon Social Learning Center, Eugene, Oregon, USA
Chapter 9.5.4

Martine F. Flament Chargée de Récherche INSERM, CNRS UMR 7593, Paris, France
Chapter 9.2.8

Simon Fleminger Consultant Neuropsychiatries, Lishman Brain Injury Unit, Maudsley Hospital, London, UK
Chapters 4.1.10 and 4.1.13

Jonathan Flint Wellcome Trust Centre for Human Genetics Roosevelt Drive, Oxford
Chapter 2.4.2

Susan Folstein Professor of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, USA
Chapter 4.1.7

Peter Fonagy Freud Memorial Professor of Psychoanalysis, University College London; Director of Research, Anna Freud Centre, London, UK; Director, Child and Family Center and Clinical Protocols and Outcomes Center, Menninger Clinic, Topeka, Kansas, USA
Chapters 4.12.6, 6.3.5 and 9.5.2

Gillian C. Forrest Consultant Child and Adolescent Psychiatrist
Chapter 9.5.8

Leah Fostick Department of Psychiatry, Chaim Sheba Medical Centre, Tel Hashomer, Israel
Chapter 4.8

W. Fraser Division of Psychological Medicine, University of Wales College of Medicine, Cardiff, UK
Introduction to Section 10

Laura Fratiglioni Aging Research Centre, Karolinska Institute, Stockholm, Sweden
Chapter 8.3

David M. Fresco Adult Anxiety Clinic of Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

K.W.M. Fulford Professor of Philosophy and Mental Health, University of Warwick; Honorary Consultant Psychiatrist, University of Oxford, UK
Chapter 1.5.2

Robert Fung, Specialist Registrar in Psychiatry, Sheffield Care NHS Trust, UK
Chapter 4.2.3.2

Pier Maria Furlan Director of Department of Mental Health San Luigi Gonzaga Hospital - University of Torino, Italy
Chapter 5.5

Glen O. Gabbard Bessie Walker Callaway Distinguished Professor of Psychoanalysis and Education in the Kansas School of Psychiatry, Menninger Clinic, Topeka; Clinical Professor of Psychiatry of Kansas School of Medicine, Wichita, Kansas, USA
Chapter 3.3

Jean Garrabé Honorary President of L'Evolution psychiatrique, Paris
Chapter 4.3.10

M. Elena Garralda Professor of Child and Adolescent Psychiatry, Imperial College of Medicine, London, UK
Chapter 9.3.4

Ann Gath Formerly of University College London, UK
Chapter 10.8

John R. Geddes Professor of Epidemiological Psychiatry, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford
Chapters 1.10 and 4.5.1

David Gill Research Fellow, Department of Psychiatry, University of Oxford, UK
Chapters 5.2.9

Kathryn J. Gill MUHB Addictions Unit, McGill University, Montreal QC, Canada
Chapter 4.2.2.1

Julia Gledhill Clinical Research Fellow, Imperial College of Medicine, London, UK
Chapter 9.3.4

David Goldberg Director of Research and Development, Institute of Psychiatry, King's College London, UK
Chapter 7.8

Terry E. Goldberg The Zucker Hillside Hospital, Glen Oaks NY, USA
Chapter 2.5.3

Guy Goodwin Professor, University Department, Warneford Hospital, Oxford
Chapter 4.5.6

Kevin Gournay Emeritus Professor, Institute of Psychiatry, King's College London
Chapter 6.4.2

Simon G. Gowers Professor of Adolescent Psychiatry, University of Liverpool
Chapter 9.5.9

Cynthia A. Graham, Oxford Doctoral Course in Clinical Psychology Warneford Hospital, Oxford and The Kinsey Institute for Research in Sex, Gender, & Reproduction
Chapter 4.11.2

Philip Graham Emeritus Professor of Child Psychiatry, Institute of Child Health, London
Chapter 9.5.3

P.M. Grasby Senior Lecturer, MRC Cyclotron Unit, Hammersmith Hospital, London, UK
Chapter 2.3.6

Richard Green Head, Gender Identity Clinic, and Visiting Professor of Psychiatry, Imperial College of Medicine at Charing Cross Hospital, London, UK; Emeritus Professor of Psychiatry, University of California, Los Angeles, California, USA
Chapters 4.11.4 and 9.2.12

Stephen Green Clinical Professor of Psychiatry, Georgetown University School of Medicine, Washington, D.C.
Chapter 1.5.1

Michael Gunn Professor of Law and Head of Department, Department of Academic Legal Studies, Nottingham Law School, Nottingham Trent University, UK
Chapter 11.1

Lars Gustafson Professor of Geriatric Psychiatry, Lund University Hospital, Lund, Sweden
Chapter 4.1.3

Francesca Guzzetta Bologna, Italy
Chapter 4.12.4

Jennifer Gutstein Department of Child Psychiatry, College of Physicians and Surgeons, Columbia University, New York, USA
Chapter 9.2.10

Sarah Harper Oxford Institute for Aging, University of Oxford, Oxford, UK
Chapter 8.2

Rex Haigh Project Lead, Community of Communities, Centre for Quality Improvement, Royal College of Psychiatrists, London; Consultant Psychiatrist, Berkshire Healthcare NHS Foundation Trust
Chapter 6.3.9

S.A. Hales Psychiatry Fellow, Princess Margaret Hospital, University Health Network, Toronto, Canada
Chapter 5.3.6

John N. Hall Professor of Mental Health, School of Health and Social Care, Oxford Brookes University, Oxford, UK
Chapter 1.8.3

Wayne Hall Professor of Public Health Policy, University of Queensland, Herston, Australia
Chapter 4.2.3.7

Edwin Harari Consultant Psychiatrist, St Vincent's Hospital, Melbourne, Australia
Chapter 6.3.8

Sarah Harper Oxford Institute for Aging, University of Oxford, Oxford, UK
Chapter 8.2

James C. Harris Director Developmental Neuropsychiatry Clinic, Professor of Psychiatry and Behavioral Sciences, Pediatrics, and Mental Hygiene, The Johns Hopkins University School of Medicine
Chapter 9.2.1

Paul J. Harrison Clinical Reader in Psychiatry, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapter 4.3.6.2

Allison G. Harvey Department of Experimental Psychology, University of Oxford, UK
Chapter 4.6.1

Philip D. Harvey Professor of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Woodruff Memorial Building, Atlanta, GA, USA
Chapter 4.3.3

Keith Hawton Director, Centre for Suicide Research, University Department of Psychiatry, Warneford Hospital, Oxford
Chapter 4.15.4

Richard G. Heimberg Adult Anxiety Clinic of Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Scott Henderson Emeritus Professor, The Australian National University, Canberra, Australia
Chapters 2.7 and 8.3

George R. Heninger Professor, Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut, USA
Chapter 6.2.3

Leslie Hicks, University of York, UK
Chapter 9.5.6

A.J. Holland Lecturer, Department of Psychiatry, University of Cambridge, UK
Chapter 10.1

Jimmie C. Holland Wayne E. Chapman Chair in Psychiatric Oncology, Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, USA
Chapter 5.3.7

Sheila Hollins Professor of Psychiatry of Learning Disability, Department of Psychiatry and Disability, St George's Hospital Medical School, University of London, UK
Chapter 10.7

Jeremy Holmes Consultant Psychiatrist/Psychotherapist, North Devon District Hospital, Barnstaple; Senior Lecturer, University of Bristol, UK
Chapter 3.2

Suzanne Holroyd Professor, Director of Geriatric Psychiatry, Department of Psychiatry and Neurobehavioral Science, University of Virginia, Charlottesville VA
Chapter 8.5.6

Geraldine Holt Honorary Senior Lecturer in Psychiatry at the Institute of Psychiatry, King's College London, UK
Chapter 10.9

Allan House Professor of Liaison Psychiatry, University of Leeds, UK
Chapter 5.3.1

Daniel H. Hovelson The Harvard program in refugee trauma, Massachusetts general hospital, Dept of psychiatry.
Chapter 7.10.1

Jane Hubert Senior Lecturer in Social Anthropology, Department of Psychiatry and Disability, St George's Hospital Medical School, University of London, UK
Chapter 10.7

Stephen Hucker University of Toronto, Toronto, Canada
Chapter 11.8

T. A. Hughes Consultant Psychiatrist, St Mary's Hospital, Leeds, UK
Chapter 4.1.6

Masud Husain Institute of Neurology & Institute of Cognitive Neuroscience, UCL, London and National Hospital for Neurology & Neurosurgery, London
Chapter 2.5.2

Frits J. Huyse Psychiatrist, Consultant integrated care, Department of General Internal Medicine, University Medical Centre Groningen (UMCG), Groningen, The Netherlands
Chapter 5.7

Matti Iivanainen Professor, Department of Child Neurology, University of Helsinki, Finland
Chapter 10.5.3

Leslie Iversen Visting Professor, Department of Pharmacology, University of Oxford, UK
Chapter 6.2.7

Richard Ives National Children's Bureau, London, UK
Chapter 4.2.3.6

Assen Jablensky Professor of Psychiatry, University of Western Australia, Perth, Australia
Chapters 4.3.5 and 4.3.7

Robin Jacoby Clinical Reader in the Psychiatry of Old Age, University of Oxford, UK
Chapters 8.4 and 8.5.7

Claudia Jacova Assistant Professor, Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada
Chapter 8.5.1.1

David V. James Consultant Forensic Psychiatrist, North London Forensic Service and Fixated Threat Assessment Centre
Chapter 11.6

Kay Redfield Jamison Professor of Psychiatry, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
Chapter 1.1

John G.R. Jefferys Department of Neurophysiology, Division of Neuroscience, University of Birmingham, UK
Chapter 2.3.9

Andrew Johns Consultant Forensic Psychiatry and Honorary Senior Lecturer, Maudsley Hospital, London, UK
Chapter 11.3.2.

E.C. Johnstone Professor of Psychiatry and Head, Department of Psychiatry, University of Edinburgh, UK
Chapter 4.3.8

David P.H. Jones Senior Clinical Lecturer in Child Psychiatry, Park Hospital for Children, University of Oxford, UK
Chapter 9.3.3

Peter R. Joyce Professor, Department of Psychological Medicine, Christchurch School of Medicine, Christchurch, New Zealand
Chapter 4.5.4

Elizabeth Juven-Wetzler Department of Psychiatry, Chaim Sheba Medical Centre, Tel Hashomer, Israel
Chapter 4.8

Horst Kachele Universitätsklinik Psychosomatische Medizin and Psychotherapie Universitätsklinik Ulm, Germany
Chapter 6.3.5

Adam Ian Kaplin Assistant Professor, Departments of Psychiatry and Neurology, Johns Hopkins University School of Medicine, Johns Hopkins Hospital, Baltimore, MD
Chapter 1.1

Markus Kaski Director, Rinnekoti Research Foundation, Director and Chief Physician of Rinnekoti Foundation, Espoo, Finland
Chapter 10.3

Aliya Kassam Institute of Psychiatry, David Glodberg Centre, De Crespigny Park, London, UK
Chapter 1.2

Roger G. Kathol, Adjunct Professor of Internal Medicine and Psychiatry, University of Minnesota, President, Cartesian Solutions, Inc. Burnsville, MN, USA
Chapter 5.7

Paul E. Keck Jr. Lindner Center of HOPE, Mason, and Department of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, OH, USA
Chapter 4.13.1

John Kellett St George's Hospital Medical School, London, UK
Chapter 8.5.8

Keith W. Kelley Department of Animal Sciences, University of Illinois, Urbana-Champaign, USA
Chapter 2.3.10

David Kennard Chair of the UK Network of the International Society for the Psychological Treatments of the Schizophrenias and other psychoses (ISPS UK); former Head of Psychology Services, The Retreat, York, UK
Chapter 6.3.9

Ad.J.F.M. Kerkhof Professor of Clinical Psychology, Vrije Universiteit, Amsterdam, The Netherlands
Chapter 4.15.2

Otto F. Kernberg Professor of Psychiatry, Cornell University Medical College, New York; Training and Supervising Analyst, Columbia University Center for Psychoanalytic Training and Research, New York, USA
Chapter 3.1

Falk Kiefer Professor of Addiction Research, Deputy Director, Department of Addictive Behaviour and Addiction Medicine, Central Institute of Mental Health CIMH, University of Heidelberg, Mannheim, Germany
Chapter 4.2.2.3

Arthur Kleinman Presley Professor of Anthropology and Psychiatry, Harvard University; Chair, Department of Social Medicine, Harvard Medical School, Cambridge, Massachusetts, USA
Chapter 2.6.2

Ami Klin Yale University, New Haven, Connecticut, USA
Chapter 9.2.3

Kimberly Klipstein Department of Psychiatry, Mount Sinai School of Medicine, New York, USA
Chapter 4.6.4

Martin Knapp Institute of Psychiatry, King's College London; London School of Economics and Political Science, University of London, UK
Chapter 7.7

Michael D. Kopelman Professor of Neuropsychiatry at King's College London, Institute of Psychiatry, UK
Chapter 4.1.12

Malcolm Lader Emeritus Professor of Clinical Psychopharmacology, King's College London, Institute of Psychiatry, Denmark Hill, London, UK
Chapter 6.2.2

Fergus D. Law Honorary Senior Registrar and Clinical Lecturer, Psychopharmacology Unit, University of Bristol, UK
Chapters 4.2.1 and 6.2.8

Brian Lawlor Conolly Norman Professor of Old Age Psychiatry, St. James's Hospital & Trinity College, Dublin, Ireland
Chapter 8.5.2

Dusica Lecic-Tosevski Professor of Psychiatry, Institute of Mental Health, School of Medicine, University of Belgrade, Belgrade, Serbia
Chapter 4.12.3

Julian Leff Emeritus Professor, Department of Psychological Medicine, Institute of Psychiatry, King's College London, UK
Chapter 1.3.2

R.J. Levin Department of Biomedical Science, University of Sheffield, UK
Chapter 4.11.1

Rhohel Lenroot Child Psychiatry Branch, NIMH, Bethesda MD, USA
Chapter 9.1.4

Peter F. Liddle Professor of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada
Chapter 4.3.2

Felice Lieh Mak Emeritus Professor, Department of Psychiatry, University of Hong Kong, Hong Kong
Chapter 5.2.10

James Lindesay Professor of Psychiatry for the Elderly, University of Leicester, UK
Chapters 8.5.1 and 8.5.5

Jouko K. Lonnqvist Professor, National Public Health Institute, Helsinki, Finland
Chapter 4.15.1

Juan J. López-Ibor Jr. Chairman, Department of Psychiatry, San Carlos University Hospital, Complutense University, Madrid, Spain
Chapters 1.6 and 4.12.1

Mª Inés López-Ibor Madrid, Spain
Chapter 4.2.3.8

Simon Lovestone Professor of Old Age Psychiatry, NIHR Biomedical Research Centre for Mental Health, MRC Centre for Neurodegeneration Research, Departments of Psychological Medicine and Neuroscience, King's College London, Institute of Psychiatry, London, UK
Chapter 4.1.2

Leanne Magee Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Andrew McCulloch The Mental Health Foundation, London, UK
Chapter 7.1

Jane McCarthy Division of Mental Health, St George's Hospital, London, UK
Chapter 10.8

Susan L. McElroy Lindner Center of HOPE, Mason, and Department of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA
Chapter 4.13.1

Peter McGuffin Director and Professor of Psychiatric Genetics, Institute of Psychiatry, King's College London, UK
Chapter 2.4.1

I.G. McKeith Clinical Director, Institute for Ageing and Health, Newcastle University, Newcastle Upon Tyne, UK
Chapter 4.1.5

Kwame McKenzie Centre for Addictions and Mental Health, Toronto, Canada; University of Toronto, Canada; University of Central Lancashire, UK
Chapter 7.3

Declan McLoughlin Institute of Psychiatry, King's College London, UK
Chapter 6.2.10.3

Mark W. Mahowald Director, Minnesota Regional Sleep Disorders Center, Hennepin County Medical Center; Professor of Neurology, University of Minnesota Medical School, Minneapolis, Minnesota, USA
Chapter 4.14.4

Mario Maj Institute of Psychiatry, University of Naples, Italy
Chapter 4.1.9

Ulrik Fredrik Malt Professor of Psychiatry (Psychosomatic Medicine), National Hospital, University of Oslo, Norway
Chapter 5.3.8

J. John Mann Vice Chair for Research Scientific Director, Kreitchman PET Center, Columbia University and Chief, Division of Molecular Imaging & Neuropathology, New York State Psychiatric Institute, USA
Chapter 4.15.3

Karl F. Mann Professor and Chair in Addiction Research, Deputy Director Central Institute of Mental Health (CIMH), University of Heidelberg, Mannheim, Germany
Chapter 4.2.2.3

Russell L. Margolis Professor of Psychiatry and Neurology Director, Johns Hopkins Schizophrenia Program Director, Laboratory of Genetic Neurobiology Division of Neurobiology, Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, USA
Chapter 4.1.7

John C. Markowitz Associate Professor of Psychiatry, Weill Medical College of Cornell University; Director, Psychotherapy Clinic, Payne Whitney Clinic, New York Presbyterian Hospital, New York, USA
Chapter 6.3.3

John Marsden Lecturer, Institute of Psychiatry, King's College London, UK
Chapter 4.2.4

Jane Marshall Senior Lecturer in the Addictions, National Addiction Centre, Institute of Psychiatry, King's College London, UK
Chapters 4.1.11 and 4.2.2.2

Andrés Martin Professor of Child Psychiatry, Child Study Center Yale University School of Medicine, New Haven, Connecticut, USA
Chapter 9.1.3

Keith Matthews Dept of Psychiatry, University of Dundee, Dundee, UK
Chapter 6.2.10.4

Barbara Maughan MRC Child Psychiatry Unit, Institute of Psychiatry, King's College London, UK
Chapter 9.3.1

Soraya Mayet National Addiction Centre, Institute of Psychiatry, King's College London, UK
Chapter 4.2.3.1

Richard Mayou Emeritus Professor of Psychiatry, University of Oxford, UK
Chapter 5.2.1

Nick Medford Institute of Psychiatry, King's College London, UK
Chapter 4.9

David Meagher Dept of Adult Psychiatry, Midwestern Regional Hospital, Limerick, Ireland
Chapter 4.1.1

Thomas W. Meeks Division of Geriatric Psychiatry, University of California San Diego, La Jolla CA, USA
Chapter 8.5.3

Pamela S. Melding Honorary Senior Lecturer, Department of Psychological Medicine, University of Auckland, New Zealand and Consultant in Psychiatry of Old Age, Mental Health Serviced, North Shore Hospital, Waitemata District Health Board, Takapuna, North Shore City, Auckland, New Zealand
Chapter 8.7

Herbert Y. Meltzer Bixler/May/Johnaon Professor of Psychiatry, Professor of Pharmacoloqy Vanderbilt University School of Medicine, Nashville, Tennessee, USA
Chapter 6.2.5

Julien Mendlewicz Department of Psychiatry, University Clinics of Brussels, Erasme Hospital, Brussels, Belgium
Chapter 4.5.5

Andreas Meyer-Lindenberg Dept of Psychiatry, Central Institute of Mental Health, Mannheim, Germany
Chapter 2.5.3

Gillian C. Mezey Consultant and Senior Lecturer in Forensic Psychiatry, Traumatic Stress Service, St George's Hospital Medical School, London, UK
Chapters 11.12 and 11.13

R.H.S. Mindham Emeritus Professor of Psychiatry, University of Leeds, UK
Chapter 4.1.6

Andrew Mogg Institute of Psychiatry, King's College London, UK
Chapter 6.2.10.3

Richard F. Mollica Director, Harvard Program in Refugee Trauma; Associate Professor of Psychiatry, Harvard Medical School and Harvard School of Public Health, Cambridge, Massachusetts, USA
Chapter 7.10.1

Emanuel Moran Consultant Psychiatrist, Grovelands Priory Hospital, London, UK
Chapter 4.13.2

Stella Morris Dept of Psychological Medicine, Hull Royal Infirmary, Hull, UK
Chapter 5.2.6

Matt Muijen WHO Regional Office for Europe, Copenhagen, Denmark
Chapter 7.1

Paul E. Mullen Professor of Forensic Psychiatry, Monash University; Clinical Director, Victorian Institute of Forensic Mental Health, Monash University, Melbourne, Australia
Chapters 11.10, 11.11 and 11.14

Alistair Munro Emeritus Professor of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada
Chapter 4.4

Rebecca Murphy Research Psychologist, Oxford University Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapters 4.10.2 and 6.3.2.2

Lynne Murray Winnicott Research Unit, University of Reading, Reading, UK
Chapter 9.3.6

R.M. Murray Institute of Psychiatry, King's College London, UK
Chapter 4.3.6.1

Norbert Nedopil Professor of Forensic Psychiatry, Head of the Department of Forensic Psychiatry at the Psychiatric Hospital of the University of Munich, Munich, Germany
Chapter 11.3.3

Juan C. Negrete Professor and Head, Addictions Psychiatry Program, University of Toronto, Canada
Chapter 4.2.2.1

Gretchen N. Neigh Dept of Psychiatry and Behavioral Sciences, Emory University, Atlanta GA, USA
Chapter 2.3.3

Charles B. Nemeroff Reunette W. Harris Professor and Chairman, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia, USA
Chapter 2.3.3

Giles Newton-Howes Division of Neurosciences and Mental Health, Imperial College School of Medicine, London, UL
Chapter 4.12.7

Jeffrey Newcorm Mount Sinai School of Medicine, New York, USA
Chapter 4.6.4

Russell Noyes Jr. Department of Psychiatry, University of Iowa College of Medicine, Iowa City, Iowa, USA
Chapter 5.2.5

David J. Nutt Professor of Psychopharmacology and Head of Clinical Medicine, University of Bristol, UK
Chapters 4.2.1 and 6.2.8

Margaret Oates Senior Lecturer in Psychiatry, University of Nottingham, UK
Chapter 1.8.1

James R.P. Ogloff Victorian Institute of Forensic Mental Health, Thomas Embling Hospital, Fairfield VIC, Australia
Chapters 11.4 ,11.14, and 11.16

John O'Grady Knowle Hospital, Fareham, UK
Chapter 11.15

Catherine Oppenheimer Consultant Psychiatrist, Warneford Hospital, Oxford, UK
Chapter 8.5.8 and 8.6

Luca Ostacoli Liaison Psychiatry and Psychosomatic Unit, Department of Mental Health, San Luigi Gonzaga Hospital - University of Torino, Italy
Chapter 5.5

Pierre Oswald Dept of Psychiatry, ULB Erasme, Brussels, Belgium
Chapter 4.5.5

Barton W. Palmer Veterans Affairs Medical Center, University of California, San Diego CA, USA
Chapter 8.5.3

Gordon Parker Professor, University of New South Wales; and Executive Director, Black Dog Institute, Australia
Chapter 4.5.3

E.S. Paykel Emeritus Professor of Psychiatry, Department of Psychiatry, University of Cambridge, UK
Chapter 4.5.8

John B. Pearce Emeritus Professor of Child and Adolescent Psychiatry, University of Nottingham, UK
Chapters 9.4 and 9.5.1

†**R.C.A. Pearson** Department of Biomedical Science, University of Sheffield, UK
Chapter 2.3.1

Keith J. Petrie Associate Professor, School of Medicine,University of Auckland, New Zealand
Chapter 5.6

Cynthia R. Pfeffer Weill Medical College of Cornell University, New York Presbyterian Hospital-Westchester Division, White Plains, New York, USA
Chapter 9.2.10

Katharine A. Phillips Professor of Psychiatry and Human Behavior, The Warren Alpert Medical School of Brown University; Director, Body Dysmorphic Disorder Program, Butler Hospital, Providence, USA
Chapter 5.2.8

Pierre Pichot Académie Nationale de Médecine, Paris, France
Chapter 1.4

Harold Alan Pincus Columbia University, New York, USA
Chapter 1.9

Vanessa Pinfold 'Rethink', London, UK
Chapter 7.9

Daniel S. Pine Division of Intramural Research Programs, National Institutes of Health, Bethesda, USA
Chapter 9.2.6

Malcolm Pines Founding Member, Institute of Group Analysis, London, UK
Chapter 6.3.6

Harrison G. Pope Jr. Professor of Psychiatry, Harvard Medical School, Boston; Chief, Biological Psychiatry Laboratory, McClean Hospital, Belmont, Massachusetts, USA
Chapter 6.2.6

Robert M. Post Chief, Biological Psychiatry Branch, National Institute of Mental Health, Bethesda, Maryland, USA
Chapter 6.2.4

Graham E. Powell Psychology Services, Powell Campbell Edelmann, London, UK
Chapter 1.8.3

Herschel Prins Professor, Midlands Centre for Criminology and Criminal Justice, University of Loughborough, UK
Chapter 11.9

Paul Ramchandani Dept of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
Chapter 9.3.6

Shulamit Ramon Professor of Interprofessional Health and Social Studies, Anglia Polytechnic University, Cambridge, UK
Chapter 6.4.3

Beverley Raphael University of Western Sydney Medical School, Sydney NSW, Australia
Chapter 4.6.5

James Reich Clinical Professor of Psychiatry, University of California, San Francisco Medical School and Adjunct Associate Professor of Psychiatry, Stanford School of Medicine
Chapter 4.12.2

Helmut Remschmidt Director, Department of Child and Adolescent Psychiatry, Philipps Universität, Marburg, Germany
Chapter 9.2.2

Philippe Robaey Institute of Mental Health Research, Royal Ottawa Hospital, Ottawa, Canada
Chapter 9.2.8

Ian Robbins Consultant Clinical Psychologist, St George's Hospital, London, UK
Chapter 11.13

T.W. Robbins Section of Forensic Psychiatry, St George's Hospital Medical School, London, UK
Chapter 2.5.5

G.M. Rodin Professor of Psychiatry, University of Toronto, Toronto, Canada
Chapter 5.3.6

Maria A. Ron Professor of Neuropsychiatry, Institute of Neurology, University College London, UK
Chapter 5.3.2

Robin Room Professor, School of Population Health, University of Melbourne; and Director, AER Centre for Alcohol Policy Research, Turning Point Alcohol and Drug Centre, Fitzroy, Victoria, Australia
Chapter 4.2.2.6

W. Rössler Professor of Clinical Psychiatry and Psychology, University of Zürich, Switzerland
Chapter 6.4.1

James R. Rundell Department of Psychiatry and Psychology, Mayo Clinic Professor of Psychiatry, Mayo Clinic College of Medicine
Chapter 5.3.4

Gerald Russell Emeritus Professor of Psychiatry, Director of the Eating Disorders Unit, Hayes Grove Priory Hospital, Hayes, Kent, UK
Chapter 4.10.1

B.J. Sahakian Dept of Psychiatry, University of Cambridge, Cambridge, UK
Chapter 2.5.5

Diana Sanders Chartered Counselling Psychologist, working in Psychological Medicine in Oxford, UK
Chapter 6.3.1

Paramala J. Santosh Great Ormond Street Hospital for Sick Children, London, UK
Chapter 9.5.5

Benedetto Saraceno Director of Department of Mental Health and Substance Abuse, World Health Organization WHO
Chapter 1.3.1

Gauri N. Savla, Veterans Affairs Medical Center, University of California, San Diego CA, USA
Chapter 8.5.3

Carlos H. Schenck Staff Psychiatrist, Minnesota Regional Sleep Disorders Center, Hennepin County Medical Center; Associate Professor of Psychiatry, University of Minnesota Medical School, Minneapolis, Minnesota, USA
Chapter 4.14.4

John Schlapobersky Consultant Psychotherapist, Trumatic Stress Clinic Middlesex/University College Hospital, formerly also of The Medical Foundation for the Care of Victims of Torture London, UK
Chapter 6.3.6

Fabrizio Schiffano, Chair in Clinical Pharmacology and Therapeutics Associate Dean, Postgraduate Medical School, Hon Consultant Psychiatrist Addictions, University of Hertfordshire, School of Pharmacy, College Lane Campus, Hatfield, UK
Chapter 4.2.3.5

Gerd Schulte-Körne Director of the Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University of Munich, Pettenkoferstr, München/Germany
Chapter 9.2.2

J. Scott Professor of Psychological Medicine, University of Newcastle & Honorary Professor, Psychological Treatments Research, Institute of Psychiatry, London and University Department of Psychiatry, Leazes Wing, Royal Victoria Infirmary, Newcastle upon Tyne, England
Chapter 4.5.8

Stephen Scott Professor of Child Health & Behaviour, King's College London, Institute of Psychiatry, and Director of Research National Academy for Parenting Practitioners, London, UK
Chapters 9.1.1 and 9.2.5

Nicholas Seivewright Consultant Psychiatrist in Substance Misuse, Community Health Sheffield NHS Trust, Sheffield, UK
Chapter 4.2.3.2

David Shaffer Department of Child Psychiatry, College of Physicians and Surgeons, Columbia University, New York, USA
Chapter 9.2.10

Trevor Sharp Dept of Pharmacology, University of Oxford, Oxford, UK
Chapter 2.3.4

Michael Sharpe Professor of Psychological Medicine & Symptoms Research, University of Edinburgh, UK
Chapters 5.1 and 5.2.7

Jennifer Shaw Centre for Suicide Prevention, The School of Medicine, University of Manchester, UK
Chapter 11.5

Mauricio Sierra Institute of Psychiatry, King's College London, UK
Chapter 4.9

Gregory Simon Investigator, Center for Health Studies, Group Health Cooperative, Seattle, Washington, USA
Chapter 5.2.2

Andrew Sims Professor of Psychiatry, University of Leeds, UK
Chapter 1.7

Ian Sinclair Professor of Social Work, University of York, UK
Chapter 9.5.6

Mike Slade Health Service and Population Research Department and Institute of Psychiatry, King's College London, UK
Chapter 7.2

Elita Smiley Consultant Psychiatrist and Clinical Senior Lecturer, Division of Community Based Sciences, Faculty of Medicine, University of Glasgow, UK
Chapter 10.2

Wolfgang Söllner Department of Psychosomatic Medicine and Psychotherapy General Hospital Nuremberg, Prof.Ernst-Nathan-Str. 1, Nürnberg, Germany
Chapter 5.7

Daniel Souery Department of Psychiatry, University Clinics of Brussels, Erasme Hospital, Brussels, Belgium
Chapter 4.5.5

Elizabeth Spencer Senior Clinical Medical Officer, Early Intervention Service, Northern Birmingham Mental Health Trust, Birmingham, UK
Chapter 6.3.2.4

David A. Spiegel Center for Anxiety and Related Disorders at Boston University, Boston, Massachusetts, USA
Chapter 4.7.1

Costas Stefanis Honorary Professor of Psychiatry, University of Athens, Greece
Chapter 1.6

Alan Stein Royal Free and University College Medical School, University College London, and Tavistock Clinic, London, UK
Chapter 9.3.6

Jessica Stiles Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, USA
Chapter 5.3.7

William S. Stone Assistant Professor of Psychology, Director of Neuropsychology Training and Clinical Services, Department of Psychiatry, Harvard Medical School, Massachusetts Mental Health Center Public Psychiatry, Division of the Beth Israel Deaconess Medical Center, Boston, USA
Chapter 4.3.9

Gregory Stores Emeritus Professor of Developmental Neuropsychiatry, University of Oxford, UK
Chapters 4.14.1 and 9.2.9

Elizabeth A. Stormshak Assistant Professor, University of Oregon, Eugene, Oregon, USA
Chapter 9.5.4

James J. Strain Professor/Director, Behavioral Medicine and Consultation Psychiatry, Mount Sinai School of Medicine, New York, USA
Chapter 4.6.4

John Strang National Addiction Centre, Institute of Psychiatry, King's College London, UK
Chapters 4.2.3.1 and 4.2.4

J. Suckling Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Addenbrookes Hospital, Cambridge, UK
Chapters 2.3.7 and 2.3.8

Nicola Swinson Centre for Sucide Prevention, The School of Medicine, University of Manchester, UK
Chapter 11.5

Michele Tansella Professor of Psychiatry and Chairman, Department of Medicine and Public Health, Section of Psychiatry, University of Verona, Italy
Chapters 7.2 and 7.6

Mary Target Senior Lecturer in Psychoanalysis, Psychoanalysis Unit, University College London; Deputy Director of Research, Anna Freud Centre, London, UK
Chapter 9.5.2

Eric Taylor Head of Department, Child & Adolescent Psychiatry, King's College London, Institute of Psychiatry
Chapter 9.2.4

John-Paul Taylor Academic Specialist Registrar, Institute for Ageing and Health Newcastle University, Campus for Ageing and Vitality, Newcastle upon Tyne, UK
Chapter 4.1.13

Pamela J. Taylor School of Medicine, Cardiff University, Cardiff, UK
Chapter 11.17

Tatiana Taylor Dept of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
*Chapter 4.15.*4

Mary Teasdale 'Rethink', London, UK
Chapter 7.9

Anita Thapar Department of Psychological Medicine, School of Medicine, Cardiff University, UK
Chapter 2.4.1

June Thoburn Emeritus Professor of Social Work, University of East Anglia, Norwich, UK
Chapter 9.3.5

Sian Thomas Chester Young People's Centre, Shester, UK
Chapter 9.5.9

Lindsay Thomson Division of Psychiatry, University of Edinburgh, Edinburgh, UK
Chapter 11.3.1

Anne E. Thompson Emeritus Professor Child and Adolescent Psychiatry, University of Nottingham, UK
Chapter 9.4

Graham Thornicroft Professor of Community Psychiatry, Institute of Psychiatry, King's College London, UK
Chapters 1.2, 7.2 and 7.6

Adolf Tobeña Professor of Psychiatry, Director of the Dept. of Psychiatry and Forensic Medicine, Autonomous University of Barcelona, Bellaterra (Barcelona), Spain
Chapter 4.12.5

Bruce J. Tonge Head Monash University School of Psychology Psychiatry & Psychological Medicine, Monash Medical Centre, Clayton, Victoria, Australia
Chapter 10.5.1

Brian Toone Consultant, Maudsley Hospital; Honorary Senior Lecturer, Institute of Psychiatry, King's College London, UK
Chapter 5.3.3

David Trickey Leicester Royal Infirmary, Leicestershire Partnership NHS Trust, UK
Chapters 9.3.2 and 9.3.7

Paula Trzepacz Eli Lilly & Co, USA
Chapter 4.1.1

Wen-Shing Tseng Professor at Department of Psychiatry, University of Hawaii School of Medicine, USA
Chapters 4.16 and 6.5

Ming T. Tsuang Behavioral Genomics Endowed Chair and University Professor, University of California; Distinguished Professor of Psychiatry and Director, Center for Behavioral Genomics, Department of Psychiatry, University of California, San Diego, CA, USA
Chapter 4.3.9

André Tylee Director, Royal College of General Practitioners Unit for Mental Health Education in Primary Care, Institute of Psychiatry, King's College London, UK
Chapter 7.8

Amy M. Ursano Department of Psychiatry, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA
Chapter 6.3.4

Robert J. Ursano Professor and Chairman, Department of Psychiatry, Uniformed Services University of the Health Sciences, F. Edward Herbert School of Medicine, Bethesda, Maryland, USA
Chapter 6.3.4

Jim van Os Professor of Psychiatric Epidemiology, Maastricht University, Maastricht, The Netherlands and Visiting Professor of Psychiatric Epidemiology Institute of Psychiatry, London, UK
Chapter 7.3

Fred R. Volkmar Yale University, New Haven, Connecticut, USA
Chapter 9.2.3

Jane Walker Clinical Lecturer and Honorary Specialist Registrar in Liaison Psychiatry, Psychological Medicine & Symptoms Research Group, School of Molecular & Clinical Medicine, University of Edinburgh, UK
Chapter 5.1

Diane Waller Professor of Art Psychotherapy, Goldsmiths, University of London
Chapter 6.4.4

Paul Walters MRC Fellow & Specialist Psychiatrist , Programme Leader MSc in Mental Health Services Research, Section of Primary Care Mental Health, Health Service and Population Research Department, David Goldberg Centre, Institute of Psychiatry, London, UK
Chapter 7.8

John Weinmann Professor of Psychology as applied to Medicine, Institute of Psychiatry, King's College London, UK
Chapter 5.6

Myrna M. Weissman Professor of Epidemiology in Psychiatry, College of Physicians and Surgeons of Columbia University; Chief, Division of Clinical and Genetic Epidemiology, New York State Psychiatric Institute, New York, USA
Chapter 6.3.3

Sarah Welch Gloucestershire Partnership NHS Foundation Trust, UK
Chapter 4.2.3.4

Ursula Werneke Consultant Psychiatrist, Norrkoping, Sweden
Chapter 6.2.9

Simon Wessely Professor of Epidemiological and Liaison Psychiatry, Institute of Psychiatry, King's College London, UK
Chapter 5.2.7

Kay Wheat Senior Lecturer in Law, Department of Academic Legal Studies, Nottingham Law School, Nottingham Trent University, UK
Chapter 11.1

Adam R. Winstock Senior Staff Specialist, Drug Health Services, Conjoint Senior Lecturer, National Drug and Alcohol Research Centre, UNSW
Chapters 4.2.3.1 and 4.2.3.5

Sally Wooding Senior Research Fellow, SciMHA Unit, University of Western Sydney, Australia
Chapter 4.6.5

Matt Woolgar Institute of Psychiatry, King's College London, UK
Chapter 2.5.1

Miranda Wolpert Director of Child and Adolescent Mental Health Services, Evidence Based Practice Unit, University College London and Anna Freud Centre, UK
Chapter 9.5.7

Lawson Wulsin Professor of Psychiatry and Family Medicine, University of Cincinnati, OH, USA
Chapter 5.7

Richard Jed Wyatt† National Institutes of Mental Health, Bethesda, Maryland, USA
Chapter 1.1

William Yule Professor of Applied Child Psychology, Institute of Psychiatry, King's College London, UK
Chapter 2.5.1

Karl Zilles Professor, Institute of Neuroscience and Biophysics, INB-3 Research Centre, Jülich and C.&O. Vogt Institute of Brain Research, University Düsseldorf, Germany
Chapter 2.3.2

Joseph Zohar Psychiatric Medical Center, Sheba Medical Center, Tel Hashomer and Sackler School of Medicine, Tel Aviv University, Israel
Chapter 4.8

SECTION 5

Psychiatry and Medicine

This page intentionally left blank

5.1

Mind–body dualism, psychiatry, and medicine

Michael Sharpe and Jane Walker

Introduction

Patients usually attend doctors because they are concerned about symptoms. When these symptoms are associated with persistent distress or disability we refer to the patient as having an illness. When assessing the patient's illness the doctor aims to make a diagnosis, on the basis of which management can be planned and prognosis made. The diagnoses available to doctors are conventionally defined as either 'medical' or 'psychiatric'. This division of illness into two types is such an accepted feature of current medical practice that we tend to take it for granted. But is it really the best way to think about patients' illnesses and to plan their care?

In order to answer this question we will examine what is meant by 'medical' and 'psychiatric' diagnoses and the assumptions underpinning this division. The disadvantages of this dualistic approach will be considered and solutions proposed.

Diagnosis

Medical diagnosis

A medical diagnosis is a label for a condition that is: (a) conventionally treated by medical doctors and (b) listed in the classifications of medical conditions such as ICD-10. Most medical diagnoses are based on identifiable bodily pathology (abnormal structure and/or function). Therefore, to make a medical diagnosis (such as cancer) doctors will seek specific bodily symptoms before confirming the presence of bodily pathology with physical signs and biological investigations (such as X-rays).

Psychiatric diagnosis

Similarly a psychiatric diagnosis is a label for a condition that is: (a) conventionally treated by psychiatrists and (b) defined in the psychiatric diagnostic classifications of ICD and DSM. Psychiatric diagnoses are not based not on bodily pathology. They are however associated with the idea of 'psychopathology', that is proposed abnormalities of the mind. Unlike bodily pathology these abnormalities of the mind cannot be objectively identified and have to be inferred from the patient's mental symptoms and their behaviour. Investigations play little or no role in diagnosis. Psychiatric diagnoses are therefore defined on the basis of symptoms and syndromes.

When is an illness psychiatric?

Why are some illnesses regarded as 'mental' or 'psychiatric' as opposed to 'medical'? Examination of the criteria for diagnoses listed as psychiatric reveals that readily observable factors common to most 'psychiatric' illnesses are:

- an absence of known bodily pathology
- an abnormal mental state as inferred by the patient's report
- a presentation with disturbed behaviour

Mind–body dualism

The underlying assumption of this dichotomous view is that it is both valid and useful to divide human illnesses into those of the body and those of the mind.[1] This idea of mind–body dualism is commonly attributed to the writings of the philosopher Descartes. So-called Cartesian dualism has exerted a profound influence on Western medical thinking and still shapes our thinking, training, and service provision.

However, dualism is at best an oversimplification and at worst a source of serious theoretical and practical problems. It may be argued that there is no such thing as a purely 'bodily' or purely 'mental' illness and that all illnesses have mental and bodily aspects.[2] Furthermore, the assumption that bodily symptoms indicate bodily pathology and that mental symptoms indicate psychopathology gives rise to specific problems: (a) when bodily symptoms occur without bodily pathology and (b) when mental symptoms occur together with bodily pathology (see Table 5.1.1).

Bodily symptoms with no bodily pathology: somatization

When patients present with bodily symptoms and bodily pathology is confirmed they are given a medical diagnosis. When patients have bodily symptoms but there is no evidence of bodily pathology the terms 'somatization' or 'somatoform disorder' are used to describe their illness. It is unclear, however, whether these illnesses are properly regarded as 'psychiatric' or as 'medical' as they do not clearly fulfil criteria for either. One solution to this dilemma is to allocate these illnesses to psychiatry. The assumption is made that their somatic symptoms are really explained by psychopathology. The absence of mental symptoms, from which psychopathology

Table 5.1.1 Diagnoses symptoms and bodily pathology

Symptoms	Bodily pathology	Diagnosis
Bodily symptoms	Present Absent	Medical diagnosis Somatization
Mental symptoms	Present Absent	Comorbidity Psychiatric diagnosis

can be inferred, is explained by the idea that the psychopathology is hidden and 'converted' into bodily symptoms by a process called 'somatization' (literally making the mental somatic). Clearly these are questionable assumptions.[3]

A second solution is to assume that the patients really do have bodily pathology in some form (even though is it unknown) and to give them a medical diagnosis of a so-called 'functional disorder' such as fibromyalgia.[4] As with somatization this approach is based on questionable assumptions.

A third, and all too common solution, is for the patient to be rejected as 'not really ill' by both psychiatry and medicine. They then end up in a no-man's land between specialities. The inadequacy of all three solutions has been particularly well illustrated by the controversy and conflict surrounding the condition called Chronic Fatigue Syndrome (CFS) or Myalgic Encephalomyelitis (ME).[5]

Mental symptoms and bodily pathology: comorbidity

When a patient has both bodily pathology and mental symptoms they are given both a medical diagnosis (based on the bodily pathology) and a psychiatric diagnosis (based on presumed psychopathology). This idea of 'comorbidity' gives rise to both theoretical and practical problems, however.

The theoretical problem concerns the psychiatric diagnosis. To make this diagnosis the doctor must identify symptoms, which are considered to be evidence of psychopathology. However, some symptoms may be considered as evidence of both psychopathology and bodily pathology. For example, a patient has mental symptoms of low mood and worthlessness, and a medical diagnosis of cancer, based on bodily pathology. Should the patient's weight loss be counted toward a psychiatric diagnosis of depression or regarded as a symptom of his cancer? There is no generally agreed answer to this conundrum (although a variety of ways of addressing it have been proposed[6]), probably because it is a manifestation of the fundamentally flawed dualistic assumption.

The main practical problem that results from making two diagnoses is a failure to adequately treat the patient. Depressive disorder comorbid with a chronic medical condition is a major cause of morbidity. However, the patient's need for psychiatric treatment often goes unmet[7] because the patient is considered to have two illnesses, each requiring diagnosis and treatment by a different speciality and the treatment of the medical condition takes precedence.

Solutions to dualism

Theoretical solutions

New scientific knowledge, such as the demonstration of a bodily (neural) basis to many 'mental' symptoms is increasingly rendering crude dualistic thinking theoretically untenable.[8] Mind and brain are coming to be regarded as two sides of the same coin—the mind/brain. This paradigm shift implies that 'psychiatric' illnesses are no more distinct from 'medical conditions' than the nervous system is separate from the rest of the body. Hence, there is a need for psychiatry to become less 'brain-less' and for medicine to become less 'mind-less'.[9] According to this new way of thinking, all symptoms, whether previously regarded as 'bodily' or 'mental' are in fact products of the mind/brain's integration of bodily, psychological, and social information. Therefore to speak of 'medical' and 'psychiatric' symptoms makes no sense. Symptoms are just symptoms.

If this paradigm shift is to be fully translated into clinical practice a new unified classification system is needed that would be used by both medicine and psychiatry. One way of achieving this might be to create a multi-axial system as is currently used by DSM-IV. However, rather than using separate axes for psychiatric and medical diagnoses, separate axes would be used for symptoms (not distinguishing between medical and psychiatric) and bodily pathology.[10] Other axes could be added to ensure that other important information is included. An example is shown in Table 5.1.2.

Practical solutions

For the present we must accept that dualism continues to shape our every day thinking, practice, and service organization. It is important therefore, that the psychiatrist is aware of the practical problems that result and is equipped with ways of addressing them. In this regard the psychiatrist is especially well placed to make a major contribution to the care of all patients by ensuring that biological, psychological, and social aspects of illness are considered in every case. This so-called 'biopsychosocial' formulation was first proposed by Engel.[11] A further enhancement of this formulation is to divide the aetiological factors into those that predisposed the patient to the illness, those that precipitated or triggered it, and those that are perpetuating it. The last group of causes is a target for treatment and the first two for prevention. A useful diagram that lists factors to consider in a biopsychosocial formulation is shown in Table 5.1.3.

Service solutions

Finally, the consequence of the professional and organizational separation of medicine and psychiatry has been a major obstacle to the integrated care of patients, especially those with comorbidity and somatoform disorders. One service solution has been the establishment of so-called liaison (linking) psychiatry services to general hospital inpatient units. Another is the increasing integration of psychological management into chronic illness management programmes.[12] However truly integrated care remains the exception rather than the rule.

Table 5.1.2 A proposed multi-axial diagnostic system for use by both psychiatry and medicine

Axis 1	Symptoms or syndrome, e.g. chronic fatigue or depression
Axis 2	Bodily pathology, e.g. cancer
Axis 3	Biological factors, e.g. autonomic arousal
Axis 4	Psychological factors, e.g. beliefs
Axis 5	Social and situational factors, e.g. bereavement

Table 5.1.3 A biopsychosocial formulation

Main factors	Subfactors	Predisposing	Precipitating	Perpetuating
Biological	Disease physiology			
Psychological	Cognition mood behaviour			
Social	Interpersonal social and occupational health care system			

Conclusion

It has been taken for granted that it is appropriate and desirable to separate patients' illnesses into medical and psychiatric types. Such an approach has had advantages in allowing specialization of training and service planning but has also created obstacles to effective patient care. It is important that practising psychiatrists are aware of these obstacles and ways of overcoming them. It also seems increasingly likely that in time better understanding of neuroscience will make dualism increasingly theoretically untenable and that a better understanding of chronic illness management will make it practically redundant. Only then will psychiatry become fully reintegrated with the rest of medicine.

Further information

White, P.D. (2005). *Biopsychosocial medicine*. Oxford University Press, Oxford.

Damasio, A.R. (1994) *Descartes' error*. GP Putnam's Sons, New York.

References

1. Miresco, M.J. and Kirmayer, L.J. (2006). The persistence of mind-brain dualism in psychiatric reasoning about clinical scenarios. *The American Journal of Psychiatry*, **163**, 913–8.
2. Wade, D.T. and Halligan, P.W. (2004). Do biomedical models of illness make for good healthcare systems? *British Medical Journal*, **329**, 1398–401.
3. DeGucht, V. and Fischler, B. (2002). Somatization: a critical review of conceptual and methodological issues. *Psychosomatics*, **43**, 1–9.
4. Wessely, S., Nimnuan, C., and Sharpe M. (1999). Functional somatic syndromes: one or many? *Lancet*, **354**, 936–9.
5. Sharpe, M. (2002). The English Chief Medical Officer's Working Parties' report on the management of CFS/ME: significant breakthrough or unsatisfactory compromise? *Journal of Psychosomatic Research*, **52**, 437–8.
6. Cohen-Cole, S.A., Brown, F.W., and McDaniel, J.S. (1993). Diagnostic assessment of depression in the medically ill. In *Psychiatric care of the medical patient* (eds. A. Stoudemire and B. Fogel), pp. 53–70. Oxford University Press, New York.
7. Moussavi, S., Chatterji, S., Verdes, E., *et al*. (2007). Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet*, **370**, 851–8.
8. Kendler, K.S. (2001). A psychiatric dialogue on the mind-body problem. *The American Journal of Psychiatry*, **158**, 989–1000.
9. Eisenberg, L. (1986). Mindless and brainless in psychiatry. *British Journal of Psychiatry*, **148**, 497–508.
10. Sharpe, M., Mayou, R., and Walker, J. (2006). Bodily symptoms: new approaches to classification. *Journal of Psychosomatic Research*, **60**, 353–6.
11. Engel, G.L. (1977). The need for a new medical model: a challenge for biomedicine. *Science*, **196**, 129–96.
12. Von Korff, M., Glasgow, R.E., and Sharpe, M. (2002). Organising care for chronic illness. *British Medical Journal*, **325**, 92–4.

5.2

Somatoform disorders and other causes of medically unexplained symptoms

Contents

5.2.1 Somatoform disorders and functional symptoms

Richard Mayou

Non-specific symptoms that are not explained by organic pathology are extremely frequent in the general population[1] and in all medical settings. Most are transient, but a substantial minority is persistent, disabling, and often associated with frequent consultation. They are likely, especially when there are multiple unexplained symptoms, to be associated with psychiatric disorder (see Chapter 5.2.3). They are widely regarded as difficult to treat but only a very small proportion is seen by psychiatrists and psychologists.

This chapter covers general issues relating to functional symptoms and syndromes and their psychiatric associations. The following chapters provide more detail about the more specific forms of somatoform disorder and about functional syndromes (pain, chronic fatigue).

Terminology of functional symptoms

The terminology is unsatisfactory.[2] These symptoms are often referred to as 'medically unexplained symptoms'. This usage has the advantage of describing the clinical problem without assumptions of aetiology, but it is unsatisfactory in that it wrongly implies that there is no medical explanation. Other generally used terms include somatization, somatoform symptoms, and functional overlay. It is perhaps most satisfactory to refer to functional symptoms and functional syndromes.

This chapter is concerned with functional symptoms whether or not they are associated with psychiatric disorder.

Aetiology

A traditional Western dualist view of aetiology as being either physical or psychological, continues to influence clinical practice and current psychiatric classifications (see Chapter 5.1). In western countries, this view has resulted in great problems in psychiatric and lay understanding, in taxonomy, and in the treatment of 'unexplained' symptoms. It has also caused bewilderment in cultures that do not share this dualist approach.

An increasingly widely held alternative view, for which there is compelling evidence, is that functional symptoms result from the interaction of physiological, pathological, and psychosocial variables.[2] A primary bodily sensation or concern (Table 5.2.1.1) is then attributed or interpreted as being of sinister significance with resulting subjective symptoms, disability, and behavioural and emotional consequences. For example, awareness of normal heart rate increase due to excitement or anxiety can result in, on the one hand, panic and, on the other, worry about heart disease, restriction of daily activities, and repeated consultation to seek investigation

Table 5.2.1.1 Causes of bodily sensations

Major pathology
Minor pathology
Physiological processes, for example:
Sinus tachycardia and benign minor arrhythmias
Effects of fatigue
Hangover
Effects of overeating
Effects of prolonged inactivity
Autonomic effects of anxiety
Lack of sleep

and reassurance. The role of these factors may vary over time during the course of any individual clinical problem.

There is considerable evidence on the ways in which psychological processes affect the interpretation of physical symptoms, whatever the underlying (major or minor) pathology or physiological processes. Cognitive-behavioural formulations emphasize the central significance of health anxiety and suggest that feedback of the physiological, cognitive, affective, and behavioural consequences of this anxiety can reinforce the physical symptoms as well as their effects on everyday life.

The process of interpretation of a bodily sensation or fear is affected by several sets of factors:

- the individual's medical experience and beliefs
- social circumstances (Table 5.2.1.2)
- personality and mental state

Once symptoms have developed they may be maintained by behavioural and psychological factors and also by the reactions of others. As with other forms of anxiety, neurobiological mechanisms may perpetuate and complicate the initial presentation.

Simple reassurance is often ineffective especially in those who, by reason of personality, are inclined to worry about their health. Misconceptions are frequently reinforced and maintained by the lack of any medical explanation for worrying symptoms or by ambiguous or contradictory advice.

The association with psychiatric disorder

The majority of functional symptoms in general populations are short lived and not associated with psychiatric disorder. There is now considerable evidence both from smaller local studies and international collaborative research that the more severe and disabling functional symptoms are associated with anxiety and depressive disorder, and that this relationship is strongest for those who have the greatest number of 'unexplained' symptoms. This is so for all ethnic groups and cultures studied.[1] There are also associations with the somatoform disorders as described below.

Table 5.2.1.2 Illness experience, which may affect the interpretation of bodily sensations and concern

Childhood illness
Family illness and consultation in childhood
Childhood consultation and school absence
Physical illness in adult life
Experience and satisfaction with medical consultation
Illness in family and friends
Publicity in television, newspapers, etc.
Knowledge of illness and its treatment

Classification of unexplained symptoms

The classification of persistent and disabling functional symptoms has taken two parallel approaches.

(a) Medical descriptive syndromes

These are very numerous, clinical patterns and terms overlap and some include assumptions about aetiology. There are cultural differences in the definition and naming. There is little evidence for the validity of separate syndromes. Lay pressure groups have increasingly claimed specific syndromes, such as alleged sensitivity to dental amalgam and many 'food allergies', which are more likely to be due to their own predicaments and the apparent lack of success of conventional medicine.[3] A small number of syndromes have now received operational diagnostic criteria which have proved valuable in clinical understanding and in planning treatment, for example the criteria for chronic fatigue (Chapter 5.2.7).

(b) Psychiatric classification

This covers both well established categories, such as anxiety and depressive disorders, and the new concept, first introduced in DSM-III, of somatoform disorder.

(i) Somatoform disorder

Somatoform disorders (Table 5.2.1.3) were seen as speculative and provisional in DSM-III. The defining feature was '*physical symptoms suggesting a physical disorder for which there are no demonstrable organic findings on known physiological mechanisms, and for which there is strong evidence, or a strong presumption, that the symptoms are linked to psychological factors or conflicts*'.

The original DSM-III classification was relatively narrow but subsequent revisions of DSM and ICD-10 have incorporated non-specific categories which have turned out to be much more prevalent in all settings.

Table 5.2.1.3 Categories of somatoform disorders in ICD-10 and DSM-IV

ICD-10	DSM-IV
Somatization disorder	Somatization disorder
Undifferentiated somatoform disorder	Undifferentiated somatoform disorder
Hypochondriacal disorder	Hypochondriasis
Somatoform autonomic dysfunction	—
Persistent pain disorder	Pain disorder associated with psychological factors (and a general medical condition)
Other somatoform disorders	Somatoform disorders not otherwise specified
—	Body dysmorphic disorder
—	Conversion disorders
Neurasthenia	—

It is important to recall that somatoform disorder remains a provisional grouping for statistical purposes rather than a grouping of categories that satisfy the normal requirements of disease entities. It nevertheless indicates a substantial clinical problem associated with considerable use of health care provisions.[3]

(ii) Factitious disorder

DSM-III also introduced another new category of 'factitious disorder' for self-inflicted physical problems. These are described in Chapter 5.2.9 and should be distinguished from deliberate falsification for external gain—*malingering*. It must be remembered that patients with factitious disorder may also suffer from unexplained symptoms attributable to somatoform or other psychiatric disorders and, indeed, not uncommonly also report symptoms of undoubted physical illness.

Somatoform disorders in DSM and ICD

There are substantial differences between the use of subcategories in DSM and ICD.[4] Neurasthenia is included in ICD-10 but is not used in any section of DSM-IV; conversion disorder is a somatoform disorder in DSM-IV but not in ICD. Both classifications include both relatively specific categories (e.g. *somatization disorder* and *hypochondriasis*) and also several very vaguely defined non-specific categories. These latter include *Undifferentiated Somatoform Disorder*, *Somatoform Autonomic Dysfunction* (ICD-10 only), and *Other Somatoform Disorders*. Although these latter have attracted less clinical and research attention, they are by far the most common forms of somatoform disorder in all epidemiological studies. So broad are the criteria that it is possible to use these categories for almost all persistent unexplained physical symptoms. Epidemiological comparisons of ICD and DSM show that the use of their rather different criteria results in substantially different prevalences of somatoform disorder in community and primary care populations.

Problems in the definition of somatoform disorder

It is widely recognized that there are serious problems[2] in the overall concept of somatoform disorders and in the definition of subcategories:

- There is no unifying theoretical basis for the whole category; it is a disparate group of problems that are not easily fitted into other parts of the classifcations.
- Comorbidity is very common, especially with anxiety disorder, depressive disorder, and personality disorder.
- Some types of somatoform disorder could be more satisfactorily reassigned to other parts of the classification (for example hypochondriasis might be renamed health anxiety and moved to anxiety disorder).
- The definitions of the less specific categories (pain disorder, Undifferentiated Somatoform Disorder) do not include any psychological criteria. Instead they rely on the description and the number of physical symptoms, the same symptoms that are used to make accompanying Axis III diagnoses. Somatization disorder has attracted disproportionate attention, but appears to be no more than an uncommon, arbitrary and unreliable extreme of the spectrum of multiple physical symptoms.
- Criteria have little meaning for cultures that do not share the western presumption of the separation of body and mind.

It has become apparent that the present classifications have no value in guiding treatment and that they are both confusing and unpopular with patients and with those who treat them.

DSM-V and ICD-11 can be expected to make large changes which will depend on the resolution of conceptual arguments and substantial further research. It is hoped that the new classification and terminology will be more reliable and valid and also be much more meaningful and acceptable.[2]

It is likely that the more specific subcategories (hypochondriasis, body dysmorphic disorder) will be reassigned to other parts of the classification and that there will be modifications in their criteria. The greatest problems relate to the much more prevalent non-specific categories. There is a consensus that a more rational operational approach is required to categorize multiple symptoms which should result in either a renamed grouping on Axis I or more logically a transfer to Axis III. It should be possible to give much greater prominence both in criteria and in accompanying text to underlying psychological and behavioural abnormalities.

Classification in clinical practice

The following chapters in this section of the book describe syndromes that have proved to have some administrative value despite the acknowledged lack of validity. Anxiety and depression are considered fully elsewhere. The recognition of anxiety and depression is important because of the therapeutic implications.

In everyday clinical practice it is rarely necessary (or helpful) to attach a somatoform label. It is more useful to be able to provide brief descriptions of the clinical problem, since these can be used as a basis for formulating treatment:

- acute or chronic
- number of physical symptoms
- the nature and pattern of symptoms (i.e. clinical syndromes such as fatigue)
- association with anxiety disorder, depressive disorder, or other specific psychiatric disorder
- beliefs about cause

Assessment and treatment

The majority of those presenting unexplained symptoms in primary care require no more than medically appropriate assessment and reassurance (Table 5.2.1.4). The latter should convey to the patient that the symptoms are accepted as real and provide an explanation for their origin as well as answering the patient's worries. It is also necessary to discuss the results of any negative investigations fully.

Symptoms that persist or recur despite reassurance are generally regarded as difficult to treat. Continuing symptoms without any specific medical explanation are likely to confirm and maintain worries about serious illness, which may be further exacerbated by secondary anxiety and behavioural consequences. Therefore

Table 5.2.1.4 General principles of assessment

Consider psychological factors from the outset
Use appropriate physical investigation to exclude physical cause
Clarify psychological and physical complaints
Clarify previous personality and concerns about physical illness
Understand patient's beliefs and expectations
Identify depression or other psychiatric disorder
Identify psychosocial problems

effective treatment depends upon sympathetic treatment that meets the needs of both the patient and the family. A multi-causal view of aetiology leads to conclusions about treatment and avoids psychiatric diagnoses that may be unacceptable to the patient. Much can be done by general practitioners or non-specialists.

The general principles of treatment (Table 5.2.1.5) are similar for all forms of unexplained symptoms, single or multiple, but individual treatment plans must take account of psychiatric diagnoses of anxiety or depression and the particular type of physical symptoms.

The treatments for particular forms of somatoform disorder are discussed in later chapters. It is important to be aware that the commonest type of somatoform disorder, Undifferentiated Somatoform Disorder is not discussed separately; treatment follows the general principles described in this and other chapters. The treatments of other functional syndromes such as irritable bowel syndrome, chronic fatigue, and atypical facial pain all depend on the therapist being familiar with these syndromes and being able to provide an appropriate combination of treatment methods. For example, the management of physical de-conditioning is central to the treatment of chronic fatigue, whereas antidepressant medication has a major role in the treatment of atypical facial pain. The chapter on chronic fatigue (Chapter 5.2.7) is an example of a functional syndrome.

Much can be achieved by components of good non-specialist care, such as the following:

- discussion and explanation of the aetiology
- treatment of any minor underlying physical problem
- anxiety management (including tapes and handouts)
- advice on diary monitoring and graded return to full activities
- specific self-help programmes (e.g. chronic fatigue, irritable bowel syndrome)
- including relatives in the assessment, discussion of the nature of the problems, and explanations of the treatment

Table 5.2.1.5 General principles of treatment

Emphasize that symptoms are real and familiar and that medical care is appropriate
Minimize and control physical care
Offer an explanation and discuss
Allow patients and families to ask questions
Discuss the role of psychological factors in all medical care
Treat any primary psychiatric disorder
Agree a treatment plan

However, chronic and recurrent problems may need specialist treatment:

- psychotropic medication (antidepressants, anxiolytics)
- cognitive behavioural therapy
- interpretative psychotherapy (individual and group)
- specific psychiatric treatment for associated psychiatric and social problems
- programme to co-ordinate and control all medical care

There is a lot of evidence on the effectiveness of a range of treatments in specialist care,[5,6] but there is much less evidence about simple routine measures. The outlook for simpler syndromes of relatively recent onset is good, but the prognosis for very prolonged chronic, multiple, or recurrent syndromes (e.g. somatization disorder) is not as good. In these circumstances the control of medical care and the prevention of further iatrogenic disability may be more realistic than cure.

Further information

Mayou, R., Kirmayer, L.J., Simon, G., *et al.* (2005). Somatoform disorders: time for a new approach. *American Journal of Psychiatry*, **162**, 847–55.

Maj, M., Akiskal, H.S., Mezzich, J., *et al.* (eds.) *Somatoform disorders*. Wiley, Chichester.

References

1. Üstün, T.B. and Sartorius, N. (1995). *World health organization. Mental illness in general health care. An international study*. Wiley, Chichester.
2. Mayou, R., Kirmayer, L.J., Simon, G., *et al.* (2005). Somatoform disorders: time for a new approach. *American Journal of Psychiatry*, **162**, 847–55.
3. Barsky, A.J., Orav, E.J., and Bates, D.W. (2005). Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity *Archives of General Psychiatry*, **62**, 903–10.
4. Fink, P., Hansen, M.S., and Oxhoj, M.L. (2004). The prevalence of somatoform disorders among internal medicine patients. *Journal of Psychosomatic Research*, **56**, 413–18.
5. Kroenke, K. and Swindle, R. (2000). Cognitive-behavioral therapy for somatization and symptom syndromes; a critical review of controlled clinical trials. *Psychotherapy and Psychosomatics*, **69**, 205–15.
6. O'Malley, P.G., Jackson, J.L., Santoro, J., *et al.* (2005) Antidepressant therapy for unexplained symptoms and symptom syndromes. *The Journal of Family Practice*, **48**, 980–90.

5.2.2 Epidemiology of somatoform disorders and other causes of unexplained medical symptoms

Gregory Simon

While nearly every psychiatric syndrome may include some somatic signs or symptoms, a specific group of syndromes has been traditionally defined as somatoform. This group of disorders is distinguished by certain key features: prominent reporting of somatic

symptoms, concern about medical illness, and frequent presentation to general medical providers. As in other categories of mental disorder, the boundaries between individual syndromes are more distinct in our systems of classification than they are in nature. Understanding that various somatoform disorders often overlap, this review is organized according to the major categories of somatoform disorder described in the ICD and DSM classification systems.

Somatization disorders

Phenomenology

The term somatization has been used to refer to a variety of clinical phenomena. One traditional view defines somatization as an inability or unwillingness to express emotional distress,[1] so that somatic symptoms are an alternative 'idiom of distress'. An alternative view defines somatization as the presentation of somatic complaints to medical providers in the presence of an occult anxiety or depressive disorder.[2] A third view defines somatization as somatic symptoms, which have no clear medical explanation.[3] While these definitions appear closely related, they identify somewhat different groups of patients. The third definition (presentation of unexplained somatic symptoms) is used by official systems of classification and by most epidemiological studies, so this review will focus on that phenomenon.

Both the ICD and DSM classification systems define somatization disorder as a chronic condition characterized by the reporting of numerous unexplained somatic symptoms.[4, 5] Recent versions of both classification systems identify a core syndrome of somatization (a persistent tendency to report multiple unexplained somatic symptoms) using a simplified set of diagnostic criteria.

Prevalence

The reported prevalence of well-defined somatization disorder appears to depend significantly on the method used for assessment. Community and primary care surveys have typically relied on structured interviews to assess the lifetime prevalence of unexplained somatic symptoms. Community surveys in North America[6] and Western Europe[7, 8] have found prevalence rates of less than 2 per cent with primary care surveys finding only slightly higher prevalence rates.[9] Data from the World Health Organization (**WHO**) multicentre primary care survey indicate that recall during structured interviews may significantly underestimate the lifetime prevalence of somatization symptoms.[10] More accurate recall of lifetime symptoms (by either repeated assessments or the use of medical records) might yield significantly higher prevalence rates.

Correlates

The prevalence of somatization disorder and unexplained somatic symptoms is typically twice as high in women as in men,[11, 12] and this difference appears at time of menarche.[13] Community and primary care surveys demonstrate a substantial overlap between somatization disorder and anxiety and depressive disorders.[7, 14, 15] Anxiety and depressive disorders also predict the subsequent onset of somatization disorder.[16]

Available data show a mixed picture regarding cross-national or cross-cultural differences in the prevalence of somatization. Studies of clinical samples find that somatic symptoms are a common accompaniment of depressive and anxiety disorders worldwide.[17–19] The WHO primary care survey documented large differences in the prevalence of unexplained somatic symptoms with a markedly higher prevalence in South America than in Europe or the United States.[9] That same study, however, found that the association between unexplained symptoms and symptoms of depression or anxiety was similar across a wide range of cultures and levels of economic development.[14] One explanation for these apparently disparate findings is that the prevalence of unexplained somatic symptoms (like the prevalence of anxiety or depressive disorder) varies widely across nations and cultures, but the association between somatic and psychological distress is universal. Countries or cultures with higher rates of anxiety or depressive disorders would be expected to have higher prevalence of somatization disorder and other somatization syndromes. Given the consistent overlap between somatization disorders and other common mental disorders, some have questioned whether these conditions actually belong in a distinct category.[20, 21]

Controversies and questions

Available data do not support a specific diagnostic threshold based on the number or distribution of unexplained somatic symptoms. An increasing number of somatic symptoms is consistently associated with increases in comorbid mood or anxiety disorder, functional impairment, and use of health services.[14,15] Mindful of this continuum, both Escobar *et al.*[22] and Kroenke *et al.*[23] have described less restrictive somatization syndromes, which, despite their higher prevalence, are strongly associated with impairment and the use of health services. Both the ICD and DSM classification systems describe subthreshold or less extreme forms of this condition characterized by a smaller number of medically unexplained symptoms.[5, 24]

Longitudinal data raise questions about the presumed stability or chronicity of somatization disorder or medically unexplained somatic symptoms. Traditional descriptions of somatization disorder emphasize its stability and chronicity. Data from the WHO primary care survey, however, suggest that individual somatization symptoms vary considerably over time.[9] While the syndrome of somatization seemed somewhat more stable than anxiety and depressive disorders (typically regarded as episodic), only half of the primary care patients, satisfying Escobar's criteria for somatization syndrome at the baseline assessment, continued to meet the criteria one year later.

Hypochondriacal disorders

Phenomenology

Both the ICD and DSM classification systems define hypochondriasis by the triad of disease conviction, functional impairment, and refusal to accept appropriate reassurance.

Prevalence

Attempts to estimate the prevalence of hypochondriasis have been limited by the absence of proven standardized methods for standardized assessment. Community surveys find prevalence rates of 1 per cent or less,[25] while primary care surveys typically find rates of approximately 5 per cent,[26, 27] while the WHO multicentre primary care survey[28] found an overall prevalence of only 0.8 per cent.

In reviewing data from the WHO survey, Gureje *et al.*[28] found that a less restrictive definition more than doubled the prevalence rate (to 2.2 per cent). Cases added by this relaxed definition did not differ significantly from those satisfying CIDI/ICD criteria, suggesting that CIDI/ICD criteria may be somewhat too restrictive.

Correlates

Despite the variation in prevalence, primary care surveys yield similar results regarding demographic correlates of hypochondriasis. The prevalence of hypochondriasis is 1.5 to 2 times as great in women as men but does not appear to vary significantly with age.[28]

Controversies and questions

While the ICD and DSM classification systems suggest that hypochondriasis is distinct from anxiety and depressive disorders, available data suggests considerable overlap. In every sample examined, hypochondriasis is strongly associated with major depression, panic disorder, and generalized anxiety disorder.[26,28–30] Among those with hypochondriasis, clinical features do not clearly distinguish those with and without a comorbid psychiatric diagnosis.[30] In addition, changes over time in anxiety or depression are consistently associated with parallel changes in symptoms of hypochondriasis.[31] As with somatization disorders, some have recently argued that hypochondriasis be re-classified as a form of anxiety disorder.[20,21]

Pain syndromes

Phenomenology

While the ICD and DSM classification systems both define somatoform pain disorders, the two systems differ in their descriptions of the clinical features. In the ICD diagnostic system, somatoform pain disorder is defined as persistent pain without clear medical explanation.[5] The DSM system[4] specifies that 'psychological factors are judged to have an important role in the onset, severity, exacerbation, or maintenance of the pain'. Both definitions are somewhat problematic. Basic research on neural changes associated with persistent pain raise doubts about the distinction between pain with and without a biomedical explanation.[32] As discussed below, longitudinal research supports the view that persistent pain causes psychological disorder as much as it supports the DSM view that pain results from psychological distress. Recent epidemiological research has attempted to avoid questions of aetiology and has examined the prevalence and correlates of persistent pain.

Prevalence

Epidemiological studies consistently find that pain syndromes are among the most common problems presented to general medical providers. Population surveys indicate that over 25 per cent of community residents suffer from recurrent or persistent pain symptoms and that 2 to 3 per cent experience disabling pain syndromes.[33] The recent WHO primary care survey[34] found that approximately 20 per cent of primary care patients suffered from persistent pain (one or more pain symptoms present for most of the last 6 months). Pain syndromes are approximately twice as prevalent in women as in men.[15]

The limited data available do not allow definite conclusions about cross-cultural or cross-national variability in pain syndromes. Some studies have documented cross-national or cross-cultural differences based on small samples of patients treated for pain syndromes—often in specialist pain clinics.[35] The WHO primary care survey[34] included both the largest number of patients with pain syndromes as well as the broadest range of cultures and levels of economic development. In that study, both the prevalence and correlates of persistent pain varied widely across sites. No clear pattern (e.g. a higher prevalence in developing or non-Western countries) was evident.

Correlates

All available data indicate that pain symptoms are strongly associated with anxiety and depressive disorders. This relationship has been consistently demonstrated in both community[36] and primary care[34] studies across a broad range of cultural and socioeconomic divides. Psychological distress is most strongly associated with pain occurring at multiple sites and pain associated with functional impairment.[37, 38] While epidemiological studies strongly support an association between pain complaints and psychological distress, this does not necessarily imply that pain is a consequence of psychological distress. Some studies find that the presence of psychological distress predicts the onset of pain syndromes,[39,40] while others support the opposite relationship—that persistent pain predicts subsequent psychological disorder.[41]

Other somatoform conditions

Specific somatoform syndromes

A number of specific somatic syndromes have been described over the last several decades. These specific syndromes are sometimes defined by the particular somatic symptoms experienced (e.g. fibromyalgia, irritable bowel syndrome, chronic fatigue syndrome) and sometimes by particular beliefs about aetiology (e.g. multiple chemical sensitivity, systemic candidiasis, electrical allergy). In every case, controversy persists about whether the somatic symptoms should be considered 'medically unexplained' (that is to say a somatoform disorder). Community surveys suggest that non-specific symptoms (such as fatigue or diffuse musculoskeletal pain) are common, but that the prevalence of strictly defined syndromes (such as fibromyalgia or chronic fatigue syndrome) varies considerably with the criteria applied.[42, 43] Most of these syndromes appear more often in women than in men.[44] Both community[43, 44] and primary care surveys[45, 46] have found several of these syndromes to be associated with anxiety and depressive symptoms. However, two studies of chronic fatigue[47] and irritable bowel[48] symptoms found that psychological distress was associated with seeking care for somatic symptoms rather than the presence or nature of the somatic symptoms themselves.

Body dysmorphic disorder or monosymptomatic hypochondriasis

Body dysmorphic disorder has recently been identified as a distinct clinical entity. The limited epidemiological data available suggest significant overlap with anxiety and depressive disorders.[8, 50] Prevalence estimates range from less than 1 per cent among unselected primary care patients,[8] to as high as 5 to 10 per cent among patients seeking cosmetic surgery[49] or outpatients with anxiety

Table 5.2.2.1 Summary of prevalence rates of specific somatoform disorders in community and primary care studies

	Community studies	Primary care studies	Notes
Somatization disorder	1%	1–2%	Diagnostic interviews probably under-count past symptoms, records reviews tend to find higher rates
Multisomatoform disorder or subthreshold somatization	5%	8–10%	Less extreme form of somatization, but still strongly associated with disability and symptoms of depression and anxiety
Hypochondriasis	1%	1–5%	Strongly associated with anxiety disorders
Persistent pain syndromes	2–3%	15–20%	Bi-directional relationship with depressive and anxiety disorders
Body dysmorphic disorder	1%	1–5%	Significantly higher in certain medical settings (dermatology, cosmetic surgery) and in people with anxiety disorders

disorders.[50] Some questionnaire studies find prevalence rates as high as 5 per cent among university students.[51]

Conversion disorders

Limited epidemiological data are available concerning conversion-type somatoform disorders. As with other varieties of unexplained somatic symptoms, these disorders appear to be more common among women[52] and are associated with increased prevalence of depressive and anxiety disorders.[53] Some evidence suggests that these conditions have declined in prevalence.[52] While many sources report that these conditions are more common in non-Western or developing countries, available epidemiological data do not necessarily support this view.[9]

Factitious disorders

No systematic data are available regarding the epidemiology of factitious disorders. The available data include numerous case reports and a small number of case series—typically drawn from medical inpatient or medical specialty settings. Because the syndrome is defined by deception, it is likely that a large proportion of cases go undetected.

Further information

Creed, F. and Barsky, A. (2004). A systematic review of the epidemiology of somatisation disorder and hypochondriasis. *Journal of Psychosomatic Research*, **56**, 391–408.

Ustun, T.B. and Sartorius, N. (eds.). (1995). *Mental illness in general health care*. John Wiley & Sons, Chichester, England.

References

1. Kleinman, A. (1977). Depression, somatization and the "new cross-cultural psychiatry". *Social Science & Medicine*, **11**, 3–10.
2. Goldberg, D. and Bridges, K. (1988). Somatic presentations of psychiatric illness in primary care. *Journal of Psychosomatic Research*, **32**, 137–44.
3. Barsky, A. (1992). Amplification, somatization, and the somatoform disorders. *Psychosomatics*, **33**, 28–34.
4. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
5. World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines*. World Health Organization, Geneva.
6. Swartz, M., Blazer, D., George, L., *et al.* (1986). Somatization disorder in a community population. *American Journal of Psychiatry*, **143**, 1403–8.
7. Grabe, H., Meyer, C., Hapke, U., *et al.* (2003). Specific somatoform disorders in the general population. *Psychosomatics*, **44**, 304–11.
8. Faravelli, C., Salvatori, S., Galassi, F., *et al.* (1997). Epidemiology of somatoform disorders: a community survey in Florence. *Social Psychiatry and Psychiatric Epidemiology*, **32**, 24–9.
9. Gureje, O., Simon, G., Ustun, T., *et al.* (1997). Somatization in cross-cultural perspective: a World Health Organization study in primary care. *American Journal of Psychiatry*, **154**, 989–95.
10. Simon, G. and Gureje, O. (1999). Stability of somatization disorder and somatization symptoms among primary care patients. *Archives of General Psychiatry*, **56**, 90–5.
11. Kroenke, K. and Spitzer, R. (1998). Gender differences in the reporting of physical and somatoform symptoms. *Psychosomatcic Medicine*, **60**, 150–5.
12. Ladwig, K., Marten-Mittag, B., Erazo, N., *et al.* (2001). Identifying somatization disorder in a population-based health examination survey: psychosocial burden and gender differences. *Psychosomatics*, **42**, 511–18.
13. LeResche, L., Mancl, L., Drangsholt, M., *et al.* (2005). Relationship of pain and symptoms to pubertal development in adolescents. *Pain*, **118**, 201–9.
14. Simon, G.E., VonKorff, M., Piccinelli, M., *et al.* (1999). An international study of the relation between somatic symptoms and depression. *New England Journal of Medicine*, **341**, 1329–35.
15. Kroenke, K., Spitzer, R.L., Williams, J.B.W., *et al.* (1994). Physical symptoms in primary care: predictors of psychiatric disorders and functional impairment. *Archieves of Family Medicine*, **3**, 774–9.
16. Gureje, O. and Simon, G. (1999). The natural history of somatisation in primary care. *Psychological Medicine*, **29**, 669–76.
17. Ulasahin, A., Basoglu, M., and Paykel, E. (1994). A cross-cultural comparative study of depressive symptoms in British and Turkish clinical samples. *Social Psychiatry and Psychiatric Epidemiology*, **29**, 31–9.
18. Escobar, J., Gomez, J., and Tuason, V. (1983). Depressive phenomenology in North and South American patients. *American Journal of Psychiatry*, **140**, 47–51.
19. Ebert, D. and Martus, P. (1994). Somatization as a core symptom of melancholic type depression. Evidence from a cross-cultural study. *Journal of Affective Disorder*, **32**, 253–6.
20. Creed, F. and Barsky, A. (2004). A systematic review of the epidemiology of somatisation disorder and hypochondriasis. *Journal of Psychosomatic Research*, **56**, 391–408.
21. Mayou, R., Kirmayer, L., Simon, G., *et al.* (2005). Somatoform disorder: time for a new approach in DSM-V. *American Journal of Psychiatry*, **162**, 847–55.
22. Escobar, J.I., Burnam, M.A., Karno, M., *et al.* (1987). Somatization in the community. *Archives of General Psychiatry*, **44**, 713–18.
23. Kroenke, K., Spitzer, R., deGruy, F., *et al.* (1997). Multisomatoform disorder. An alternative to undifferentiated somatoform disorder for the somatizing patient in primary care. *Archives of General Psychiatry*, **54**, 352–8.

24. American Psychiatric Association. (1995). *Diagnostic and statistical manual of mental disorders* (DSM-IV) (4th edn), *primary care version*. American Psychiatric Press, Washington.
25. Looper, K.J. and Kirmayer L.J. (2001). Hypochondriacal concerns in a community population. *Psychological Medicine*, **31**, 577–84.
26. Escobar, J., Gara, M., Waitzkin, H., *et al.* (1998). DSM-IV hypochondriasis in primary care. *General Hospital Psychiatry*, **20**, 155–9.
27. Barsky, A., Wyshak, G., Klerman, G., *et al.* (1990). The prevalence of hypochondriasis in medical outpatients. *Social Psychiatry and Psychiatric Epidemiology*, **25**, 89–94.
28. Gureje, O., Ustun, T., and Simon, G. (1997). The syndrome of hypochondriasis: a cross-national study in primary care. *Psychological Medicine*, **27**, 1001–10.
29. Noyes, R.J. (1999). The relationship of hypochondriasis to anxiety disorders. *General Hospital Psychiatry*, **21**, 8–17.
30. Barsky, A., Wyshak, G., and Klerman, G. (1992). Psychiatric comorbidity in DSM-IIIR hypochondriasis. *Archives of General Psychiatry*, **49**, 101–8.
31. Simon, G., Gureje, O., and Fullerton, C. (2001). Course of hypochondriasis in an international primary care study. *General Hospital Psychiatry*, **23**, 51–5.
32. Coderre, T., Katz, J., Vaccarine, A.L., *et al.* (1993). Contribution of central neuroplasticity to pathological pain: review of clinical and experimental evidence. *Pain*, **52**, 259–85.
33. Von Korff, M., Dworkin, S.F., and Le Resche, L. (1990). Graded chronic pain status: an epidemiologic evaluation. *Pain*, **40(3)**, 279–91.
34. Gureje, O., Von Korff, M., Simon, G.E., *et al.* (1998). Persistent pain and well-being. A World Health Organization study in primary care. *Journal of American Medical Association*, **280**, 147–51.
35. Sanders, S., Brena, S., Spier, C., *et al.* (1992). Chronic low back pain patients around the world: cross-cultural similarities and differences. *Clinical Journal of Pain*, **8**, 317–23.
36. McWilliams, L., Cox, B., and Enns, M. (2003). Mood and anxiety disorders associated with chronic pain: an examination in a nationally representative sample. *Pain*, **106**, 127–33.
37. Dworkin, S.F., Von Korff, M., and LeResche, L. (1990). Multiple pains and psychiatric disturbance: an epidemiologic investigation. *Archives of General Psychiatry*, **47**, 239–44.
38. Benjamin, S., Morris, S., McBeth, J., *et al.* (2000). The association between chronic widespread pain and mental disorder: a population-based study. *Arthritis and Rheumatism*, **43**, 561–7.
39. Von Korff, M., Le Resche, L., and Dworkin, S.F. (1993). First onset of common pain symptoms: a prospective study of depression as a risk factor. *Pain*, **55**, 251–8.
40. Croft, P., Papegeorgiou, A., Ferry, S., *et al.* (1995). Psychologic distress and low back pain. Evidence from a prospective study in the general population. *Spine*, **15**, 2731–7.
41. VonKorff, M. and Simon, G.E. (1996). The relationship between pain and depression. *British Journal of Psychiatry*, **168**(Suppl. 30), 101–8.
42. Wessely, S., Chalder, T., Hirsch, S., *et al.* (1997). The prevalence and morbidity of chronic fatigue and chronic fatigue syndrome: a prospective primary care study. *American Journal of Public Health*, **87**, 1449–55.
43. Wolfe, F., Ross, K., Anderson,J., *et al.* (1995). The prevalence and characteristics of fibromyalgia in the general population. *Arthritis and Rheumatism*, **38**, 19–28.
44. Pawlikowska, T., Chalder, T., Hirsch, S., *et al.* (1994). Population-based study of fatigue and psychological distress. *British Medical Journal*, **308**, 763–6.
45. Wessely, S., Chalder, T., Hirsch, S., *et al.* (1996). Psychological symptoms, somatic symptoms, and psychiatric disorder in chronic fatigue and chronic fatigue syndrome: a prospective study in the primary care setting. *American Journal of Psychiatry*, **153**, 1050–9.
46. Henningsen, P., Zimmermann, T., and Sattel, H. (2003). Medically unexplained physical symptoms, anxiety, and depression: a meta-analytic review. *Psychosomatic Medicine*, **65**, 528–33.
47. Lawrie, S., Manders, D., Geddes, J., *et al.* (1997). A population-based incidence study of chronic fatigue. *Psychological Medicine*, **27**, 343–53.
48. Whitehead, W., Bosmajian, L., Zonderman, A., *et al.* (1988). Symptoms of psychologic distress associated with irritable bowel syndrome. Comparison of community and medical clinic samples. *Gastroenterology*, **95**, 709–14.
49. Sarwer, D., Wadden, T., Pertschuk, M., *et al.* (1998). Body image dissatisfaction and body dysmorphic disorder in 100 cosmetic surgery patients. *Plastic and Reconstructive Surgery*, **101**, 1644–9.
50. Simeon, D., Hollander, E., Stein, D., *et al.* (1995). Body dysmorphic disorder in the DSM-IV field trial for obesessive-compulsive disorder. *American Journal of Psychiatry*, **152**, 1207–9.
51. Bohne, A., Wilhelm, S., Keuthen, N., *et al.* (2002). Prevalence of body dysmorphic disorder in a German college student sample. *Psychiatry Research*, **109**, 101–4.
52. Singh, S. and Lee, A. (1997). Conversion disorders in Nottingham: alive, but not kicking. *Journal of Psychosomatic Research*, **43**, 425–30.
53. Sar, V., Akyuz, G., Kundakci, T., *et al.* (2004). Childhood trauma, dissociation, and psychiatric comorbidity in patients with conversion disorder. *American Journal of Psychiatry*, **161**, 2271–6.

5.2.3 Somatization disorder and related disorders

Per Fink

Introduction

The essential feature of somatization disorder and related disorders is that the patient presents multiple, medically unexplained symptoms or functional somatic symptoms. These physical complaints are not consistent with the clinical picture of known, verifiable, conventionally defined diseases, and are unsupported by clinical or paraclinical findings. The phenomenon of medically unexplained symptoms cannot simply be classified into one or a few diagnostic categories, but must be regarded as an expression of a basic mechanism by which people may respond to stressors as in the cases of depression and anxiety.[1–3] Somatization disorder and related disorders must thus be considered to possess a spectrum of severity.[3, 4] In this chapter, the focus will be on the chronic and multisymptomatic forms.

The somatization disorder diagnosis has its origin in the concept of hysteria. It was introduced in DSM-III in 1980 and was based on the criteria for 'Briquet's syndrome', a syndrome described in the early 1960s by Perley and Guze.[5] They listed 59 physical and psychological symptoms distributed in 10 groups: 25 of the symptoms from nine groups were required to qualify for the diagnosis of somatization disorder. All psychological symptoms were eliminated in the DSM-III modification to avoid overlapping with other diagnoses.

The diagnostic criteria for DSM somatization disorder varied until the introduction of the current DSM-IV. The diagnosis was included in ICD-10 in 1992, but the ICD-10 criteria list different

symptoms, and require a different number of symptoms compared with the corresponding DSM criteria.

The somatization disorder diagnosis has been criticized for being too rigid for clinical use. Only the most severe cases with a specific predefined symptom profile fulfil the diagnostic criteria, and the majority of those with multiple symptoms fall into one of the residual categories of 'undifferentiated' or 'not otherwise specified' somatoform disorders.[6]

To increase the sensitivity, Escobar *et al.*[7] introduced an abridged somatization index. This required 4 symptoms for males and 6 symptoms for females out of the 37 somatic symptoms listed in the DSM-III, compared with 12 and 14 symptoms respectively for the full DSM-III somatization disorder diagnosis. Kroenke *et al.*[8] have suggested a diagnosis of 'multisomatoform disorder', defined as three or more medically unexplained physical symptoms from a 15-symptom checklist along with at least a 2-year history of medically unexplained symptoms.

However, these abridged versions share the same basic problem as the original ones, namely that the chosen number of symptoms to qualify for the individual diagnoses is arbitrary and not empirically based. Recently a new empirically based construct was introduced, and this may have solved the problem. This new 'bodily distress disorder' diagnosis is based on positive criteria and not solely on the exclusion of all organic possibilities.[3] (Table 5.2.3.1)

This chapter will not differentiate between the different subcategories of somatoform and related disorders that are present along with somatic symptoms.

Table 5.2.3.1 Symptoms of and diagnostic criteria for bodily distress disorder

Yes	No	Symptom groups
		≥ 3 Cardiopulmonal/autonomic arousal Palpitations, heart pounding, precordial discomfort, breathlessness without exertion, hyperventilation, hot or cold sweats, trembling or shaking, dry mouth, churning in stomach, "butterflies", flushing or blushing
		≥ 3 Gastrointestinal arousal Frequent loose bowel movements, abdominal pains, feeling bloated, full of gas, distended, heavy in the stomach, regurgitations, constipation, nausea, vomiting, burning sensation in chest or epigastrium
		≥ 3 Musculoskeletal tension Pains in arms or legs, muscular aches or pains, feelings of paresis or localized weakness, back ache, pain moving from one place to another, unpleasant numbness or tingling sensations
		≥ 3 General symptoms Concentration difficulties, impairment of memory, feeling tired, headache, memory loss, dizziness
		≥ 4 symptoms from one of the above groups
Diagnostic criteria:		
1–3: 'yes': Moderate 'bodily distress disorder'		
4–5: 'yes': Severe 'bodily distress disorder'		

Clinical features

Physical symptoms and complaints

Patients with somatization and related disorders may complain of any medically unexplained non-verifiable subjective physical symptoms, and the symptoms may refer to any part or system of the body.

Complaints can be divided into:

- *Subjective symptoms*, which are sensations and other complaints that cannot be verified by another individual or by general methods of examination (e.g. pains and paraesthesia).
- *Objective symptoms*, which are complaints that can be verified if present at the time of examination (e.g. haematuria, icterus, etc.).

Findings can be divided into:

- *Provoked findings*, i.e. symptoms or signs (such as soreness resulting from pressure or sensory impairment), which the patient is unaware of until these are provoked during the physical examination
- *Certain findings*, which include objective symptoms that are verified and phenomena unnoticed by the patient but found during the physical examination (such as an abdominal tumor)

The symptom complaints in patients with somatization disorder and related disorders are dominated by subjective symptoms and provoked findings, whereas objective symptoms and positive certain findings and paraclinical findings are unusual.

Subjective symptoms may be considered to be psychological phenomena arising from personal experiences, which others cannot judge or measure, despite the fact that these symptoms could be fully explained by the presence of organic pathology. This means that there are considerable inter-individual, cultural, and historical variations in the symptom presentation, which are determined by the patient's life experience and sociocultural background, and the setting in which the patient is seen also plays a role.[9] However, the patients may also present verifiable symptoms and signs due to a physical disease or defect, which they exaggerate and incorporate into their illness. Incidental inborn errors or degenerative changes, which are asymptomatic in most individuals, may tenuously be assigned clinical importance by the doctor or the patient. For instance, degenerative changes in the spinal column are seen in most individuals when they become older, and most do not have any pain, but a patient's backache may be attributed to those changes. Furthermore, over time, the patient with a chronic condition is likely to have undergone multiple tests, invasive procedures, operations, and received medications for treatment or diagnostic purposes, and this may cause not only iatrogenic harm but also physical complications.[10] Finally, the patient may have a concurrent physical disease. The presented symptoms may thus be a difficult mixture of complaints of both organic and non-organic origin.[11]

The patients typically complain of *multiple physical symptoms*, but the number of symptoms reported by the patients vary considerably from one patient to another and over time in the same individual. The patients may complain of multiple, medically unexplained symptoms in numerous bodily systems at presentation, but sometimes the complaints are concentrated on one

symptom pattern at one time (e.g. a gastrointestinal illness) and on a different symptom pattern at another (e.g. a cardiopulmonary illness).[11] This single-organ illness picture may be due to physicians being inclined to focus their attention and investigation on the organ of their own specialty—especially in a multisymptomatic, complex patient, i.e. a gastroenterologist will focus on gastrointestinal symptoms and may ignore musculoskeletal symptoms. A new set of symptoms from another organ system may come to attention when diagnoses and treatment options have been ruled out for the current complaint. Iatrogenic factors may thus contribute significantly to the presented symptom pattern and changes in symptom pattern. Patients with somatization disorder are often inconsistent historians. They may supply incorrect information about previous episodes of their illness, minimizing or ignoring earlier instances of illness and exclusively focusing on the current symptom pattern. This may be because the patients find it difficult to account for their complicated medical history, or because they do not want to confuse the doctor with what they believe is irrelevant information. Therefore, the full clinical picture often only becomes evident after a full medical history has been obtained and the patient has been followed for some time.

Symptoms and findings are not idiosyncratic but need clarification or specification before becoming meaningful clinically. For instance, a patient complains of chest pain. There are multiple causes for chest pain, so for the doctor this is not very informative. A clarification or specification is needed. A retrosternal-localized pain of a pressing nature offers the doctor a very different diagnostic association than a chest pain that is described as being stabbing and located in the left side of the chest. In somatization disorder and related disorders, the patient usually presents a *vague illness picture* with symptoms of *non-specific* character and of low diagnostic value, i.e. symptoms that are common in the general population and which are found in many different mental and physical disorders (symptoms like fatigue, nausea, headache, dizziness).

The presentation of medically unexplained symptoms is *atypical*, that is to say the symptoms lie outside what is usual in an authentic physical disease.[11] However, the patients may have 'learned' the typical symptom presentation from different sources. For example, a patient with atypical asthma-like attacks shared a room with a patient with genuine asthma during her third hospital stay; subsequently, her attacks took on a more 'authentic' appearance.[11]

Descriptions of *symptoms are usually vague, imprecise, and inconsistent*, and the patients often have difficulties giving further details about their illness and symptoms, i.e. describing the quality, intensity, and chronology. The symptoms are described as being of maximum intensity all the time, but if the patients keep a diary of symptoms or if information is gathered from other sources like relatives or the family physician, a considerable variation in the symptom intensity from day to day or from year to year often surfaces. The patients may have difficulties in the chronology of symptoms, mixing current and past symptoms and illness episodes in a disorganized and confusing manner. It is difficult for the patients to identify relieving factors or behaviour and to identify triggering events or things that make them worse, or these are multiple, or vague and unspecific. This is in contrast to patients with physical disease, who usually describe their symptoms in a consistent and precise manner (Table 5.2.3.2).

Typically, there is a marked *discrepancy* between a patient's subjective complaints and reports on his or her functioning when this is compared with the way the patient is observed to act, move

Table 5.2.3.2 Characteristic differences in symptom description and other characteristics of well-defined somatic and related disorders including functional somatic syndromes

	Somatization disorder and related disorders	Physical disease
Symptom description		
Location	Vague, diffuse, alternating	Well-defined, constant
Intensity	Vague, indistinctly defined intensity, few variations, often at maximum at all times	Well-defined changes and levels of intensity
Periodicity	Diffuse, difficult to define, are often denied	Typically well-defined periods with aggravation or improvement
Relieving or aggravating factors	Vague, indistinct, numerous	Well-defined, few
Number of symptoms	Numerous, vague	Few, well-defined, clearly described
The nature of symptoms	Unspecific	Specific
The character of the symptoms	Uncharacteristic	Characteristic
Iatropic symptoms and main complaints	Vague, difficult to identify	Can be identified and delimited from comorbid symptoms
Description	Affective, emotional, interpreting	Clear and descriptive
Other characteristics		
Treatment and medication	Difficult to assess the effect, transitory	Level of effect well-defined
Previous treatment	Unclear what treatment the patient has undergone. Diagnostic tests are often interpreted as treatment	The patient can account for previous treatments
Emotionality	Inadequate, e.g. exaggerated or marked unaffectedness ('la belle indifférence')	Adequate - empathic

and perform during the examination, or compared with information from other sources like family. For instance, the patient moves and sits completely freely despite complaining of severe back pain or gives detailed information despite complaints of severe memory impairment.

There may be *emotional discrepancy* in which the patient shows a lack of concern about the nature and implications of the symptoms despite presenting severe symptoms that are threatening the patient's future functioning and quality of life. Other patients may in turn be very affected and emotional in their description of the symptoms and illness, describing in a colourful, dramatic, and strikingly graphic manner.

The patient's centre of attention is typically on the suffering, on the psychosocial consequences, and the restrictions that the symptoms impose on their life. On the contrary, patients with well-defined physical disease are concerned or worried about the implication of their disease for their future health, i.e. will they recover or will they die from the disease. This emotional or *psychosocial communication* among patients with somatization disorder may put pressure on the doctor to do something.

Patients with somatization disorder and related disorders usually attribute their symptoms to a physical disease, and in some cases they persistently may *refuse to accept medical reassurance* despite appropriate medical evaluations. The ICD-10 criteria for somatoform disorders include this refusal to accept medical reassurance, but recent research indicates that many patients are unsure what is wrong with them, and they do not necessarily refuse non-medical explanations if they are presented in a meaningful and understandable way.[12] Although the patients may recognize that their physical symptoms are caused by, e.g. stressful events, this does not make the symptoms disappear, and they still need treatment. The weight of the symptom (refusing to accept medical reassurance) in the medical literature therefore seems out of proportion. However, in the most severe cases, patients may be involved in patient organizations fighting for their illness to be recognized as an 'authentic' physical disease or fighting for a particular causality of their illness as, e.g. whiplash-associated disorder or hypersensitivity to electricity or chemicals (multiple chemical sensitivity). The patients may also fight for disability pension or financial compensation. This is often more a question of getting their illness recognized than receiving financial compensation. Some patients may be preoccupied with the idea that they have been mistreated or neglected by doctors, and this group will often become involved in conflicts with doctors and in legal disputes.

In some cases, there is a sudden onset of the disorder in connection with a medical condition or a trauma in a previously normal and healthy individual. It could be a whiplash trauma, a fracture, an infection, or an acute intoxication. The symptoms persist despite the original disease being cured according to a biomedical or a surgical judgement. Instead, the illness worsens and more symptoms may emerge. Our knowledge about such disorders with abrupt onset is sparse.

Psychological symptoms and comorbidity

At examination, the patients may deny emotional symptoms or conflicts, and when they do report them, they often blame them on their physical affliction. Patients may also be reluctant to display emotional difficulties because of bad experiences of doing so. They may have experienced that doctors did not believe them or accused them of making up their symptoms and have consequently felt that their physical problems are not taken seriously. However, sooner or later, most patients will exhibit emotional difficulties, and if the patients feel understood by the physician, emotional problems may as well be presented. Patients may present many different types of emotional symptoms, often unspecific, but prominent anxiety and depressive symptoms are prevalent. Although the symptoms may be as marked as in affective and anxiety disorders, they are usually more transient, changing from one day to another and especially related to specific events. At times, the psychological symptoms may fulfil the criteria for a mood or anxiety disorder; but it is characteristic that the illness picture shows variations in both bodily and emotional symptoms.

Suicidal attempts are unusual but may occur especially among severe cases, but suicide is rare. Substance abuse is frequent, whether or not this is iatrogenic sanctioned.

The way the patients present themselves is inextricably linked to personal style and possible personality disorder. As a broad spectrum of personality disorders or traits[13] is associated with somatization disorder, the presenting style varies greatly from one patient to another. Characteristically, three broad patterns of personality style may be found in these patients, especially in chronic cases: dramatic–emotional type, paranoid–hostile type, and passive–aggressive-dependent type. The same patient may show all three patterns. In less severe cases it is often observed clinically that the patients have previously been very active and hard working, have conformed socially, and had a strong social network with many responsibilities. The patients often display perfectionist traits and prefer to be in control of a situation.

Illness behaviour

Typically, patients with somatization disorder persistently exhibit consulting behaviour which results in an excessive use of medical services and alternative therapies. In chronic cases, they have often been subject to a large number of futile examinations, surgery, and medical/surgical attempts at treatment.[10,14,15] However, some patients realize quite early in the illness course that the doctors cannot help them, or they are well managed by their family physician, so they do not necessarily display this consultation behaviour.

Due to negative results of medical check-up and treatment attempts and the patients' persistent belief that they must have a physical disease, the patients may consult different physicians. The patients may have been, or may feel, mistreated or neglected by doctors and therefore want to get a second opinion, or they want to find a doctor who can help them. Sometimes this behaviour, together with the patients' personality, can result in disagreement and a mutual hostility between the patients and their doctors.

Furthermore, the different illness patterns at different times, combined with the patients' seductive, demanding personality style, may result in disagreements between the different health care professionals involved in their care, which may complicate their care.

In chronic cases, all aspects of the patients' social and family life may be centered around their illness, so that the whole of their family life is adjusted to the patients' demands ('illness as a way of living').[16]

Classification

The diagnostic criteria for somatization disorder and related disorders have varied, with different permutations of the diagnostic terminology reflecting difficulties in classification and in establishing valid criteria. The distinction between the individual somatoform disorders is unclear, which means that the majority of patients will exhibit clinical characteristics from different diagnostic categories.[6]

Except for hypochondriasis or health anxiety, the somatoform disorder categories are primarily based on the number or specificity of bodily symptoms and on the duration of illness.[17] The disorders can be divided into acute and chronic forms, into a multisymptomatic form, and into a form in which the patients only present few symptoms or symptoms mainly referring to a single organ system. The somatization disorder diagnosis includes the most chronic multisymptomatic cases lasting for 2 years or more.

The ICD-10 criteria require the following:

1 at least 2 years of multiple and variable physical symptoms for which no adequate physical explanation has been found;

2 persistent refusal to accept the advice or reassurance of several doctors that there is no physical explanation for the symptoms;

3 some degree of impairment of social and family functioning attributable to the nature of the symptoms and resulting behaviour.

The ICD-10 criteria require 'multiple physical symptoms' to include at least 6 out of 14 predefined symptoms, involving at least 2 of the following: gastrointestinal, cardiovascular, urogenital, or skin or pain symptoms. In contrast, the DSM-IV criteria demand four pain symptoms, two gastrointestinal symptoms, one sexual symptom, and one pseudoneurological symptom that are not fully explained by a medical condition. No specific symptoms are listed, but examples are given. Consequently, there is only poor to moderate agreement between the DSM-IV and ICD-10 diagnostic criteria.[16] Cases lasting less than 2 years are classified as undifferentiated somatization disorder. In ICD-10, multiple symptoms are required, which is not the case in DSM-IV.

Functional or somatoform diagnoses defined mainly by the number or specificity of bodily symptoms include, besides somatization disorder (F45.0 and 300.81), undifferentiated SD (F45.1 and 300.81), persistent somatoform pain disorder (F45.4 and 307.80), and somatoform disorder unspecified/NOS (F45.9 and 300.82). In ICD-10, this also includes somatoform autonomic dysfunction.[3–5] There may be reasons for also including neurasthenia (F48.0) into the group. The somatoform disorder concept has never been accepted among non-psychiatrists, which has led to the introduction of many different functional somatic syndromes, e.g. chronic fatigue syndrome (CFS), fibromyalgia, irritable bowel syndrome (IBS), and chronic benign pain syndrome, and new syndromes are intermittently introduced.[18]

The newly introduced diagnosis of bodily distress syndrome or disorder may be a solution to this classification problem, although it has not yet been sufficiently tested in daily clinical practice.[3] The suggested diagnosis is based on an analysis of a large sample of patients from different medical settings, and it seems to encompass the various functional syndromes advanced by different medical specialties as well as somatization disorder and related diagnoses of the psychiatric classification. The disorder may have different manifestations, i.e. GI, MS, or CP syndromes as shown in Table 5.2.3.1.

Diagnosis

A somatization disorder should be suspected in any individual with a vague or complicated medical history or unaccountable non-responsiveness to therapy. Patients with somatization disorder may not have or may deny emotional symptoms or conflicts, so the absence of significant emotional symptoms at the general psychiatric interview and history taking will not exclude the diagnosis. But the presence of a previous or current emotional disturbance does support the diagnosis, as do previous episodes of medically unexplained bodily symptoms. Taken at face value, the physical symptoms are only of modest diagnostic importance, whereas unspecific or atypical symptoms in several bodily systems, or a very unusual presentation, speak in favour of the diagnosis. Multiple fluctuating symptoms of obscure origin, and their onset before the age of 30, strongly support the diagnosis. The diagnostic criteria displayed in Table 5.2.3.1 may be used, or the criteria listed in the DSM-IV or ICD-10 diagnostic criteria for somatoform disorders may be used.

Differential diagnosis

Mental and somatoform disorders

In **malingering**, the patient feigns illness with a conscious motivation to avoid responsibility or to gain an advantage. In **factitious disorder**, the symptoms are intentionally produced and the patient may self-inflict or induce diseases and lesions. In contrast to malingering, there is no external incentive for producing the symptom(s), and the motive is unconscious and only understandable in a psychopathological context.[19, 20] In somatoform disorders, both the symptom-producing behaviour and the motive are believed to be unconscious. However, factitious or malingering symptoms, mixed with other non-intentional symptoms, may occur in somatization disorder and related disorders.[11]

Hypochondriasis or health anxiety is mainly defined in cognitive terms with the emphasis on a preoccupation with physical appearance or the fears of harbouring or developing a serious physical disease. The other categories of somatoform disorders put more emphasis on bodily symptoms. In **dissociative or conversion disorder** the patients usually present fewer symptoms, but these are almost exclusively pseudoneurological symptoms. The onset is sudden, and closely associated in time with traumatic events, insoluble and intolerable problems, or disturbed relationships. The symptoms are transient and often remit suddenly after a few days, although they may persist for longer, but seldom for more than a few months. Episodes of dissociative or conversion disorders frequently occur in patients with other somatoform disorders.

In **pain disorder**, the predominating complaints are of medically unexplained pain in one or more anatomical sites. Various aches and pains are common in somatization disorder but are more fluctuating and not so dominating in the clinical presentation since they merge with other complaints.

In **ICD-10 somatoform autonomic dysfunction**, the patients complain of symptoms associated with a specific system or organ

that is largely or completely under autonomic innervation and control. For example, the patient refers the symptoms to the heart and cardiovascular system, the gastrointestinal tract, respiratory system, genitourinary system, etc.

In most **other mental disorders** physical symptoms are prominent; it is the rule, rather than the exception, that patients with mental disorders consult their family physician because of physical and not emotional symptoms.[2, 21] The symptoms may be misinterpreted by the patient and the doctor as being caused by a physical disease. However, in these cases of 'presenting' or 'facultative' somatizing, the patient will accept the diagnosis of a psychiatric disorder when it is established and will accept that the symptoms are attributable to a psychic rather than a physical affliction.[2]

Autonomic bodily symptoms are prominent in panic disorder and generalized anxiety disorders, but the emotional component of the disorders is unmistakable and the patients will not, or only temporarily, attribute their symptoms to a physical disease.

In psychoses, particular schizophrenic physical symptoms and hypochondriacal belief are common. However, the complaints are of a paranoid quality and other psychotic symptoms are prominent, except in the case of hypochondriacal paranoia. Psychotic episodes may occur in somatization disorder, but these are usually transient.[22]

In depressive disorders, the patient's complaints are mood congruent, and the symptoms will disappear if the depression is treated. In a few cases, however, a depression may be expressed in mainly bodily complaints (i.e. masked depression), and if left untreated may result in a prolonged course.

In obsessive-compulsive disorder (OCD), the patient may fear contracting a disease from an outside source (e.g. dirt, germs, viruses, etc.). The OCD disorders share many similarities with hypochondriasis, i.e. cognitive and emotional complaints are the main problem, whereas bodily symptoms seldom are dominating.

Medical conditions

The onset of multiple physical symptoms after the age of 40, in a previously physically and mentally healthy individual, suggests a general medical condition. However, multiple vague physical complaints may be prominent in mental disorders with late onset such as mood disorders, dementia, and withdrawal states.

Only a limited number of general medical conditions present vague, non-specific, and multiple somatic symptoms (e.g. hyperparathyroidism, hyperthyroidism, acute intermittent porphyria, myasthenia gravis, AIDS, multiple sclerosis, systemic lupus erythematosus, lyme disease, and connective tissues disease). In most of these medical conditions there will be positive paraclinical findings.

In genuine physical disorders, the key symptoms will usually be characteristic from one patient to another and across illness episodes. In contrast, the constellation of symptoms in somatization disorder and related disorders will usually be incompatible with any known, authentic physical disease, and the symptom picture will be of a more fluctuating nature.

Epidemiology

The prevalence of somatization disorder is about 1 per cent in the general population and 1 to 6 per cent in primary care and in inpatient medical settings, but it is much higher if less restrictive, abridged criteria are used.[3, 7, 8, 21, 23–28] Both among primary care patients, newly referred neurological in and outpatients, and among internal medicine inpatients, the prevalence of severe bodily distress disorder is 3.3 per cent, whereas the prevalence of modest bodily distress is 25.3 per cent.[3] Regardless of the criteria used, females predominate with a male-to-female ratio of 1:2 to 1:6, and the prevalence is rather constant as to age.[21, 24–29]

The age of onset is usually before 30 to 40 years.[14] This patient group poses a considerable financial burden to health and social service provision, a loss to industry, and a dependence on invalidity benefits for long periods.[14, 24, 30, 31]

Somatization disorder and related disorders are associated with a wide spectrum of heterogeneous psychiatric disorders, including personality disorders and mental retardation. The comorbidity rates are highest in the most chronic cases, in which rates of 50 per cent or higher are reported.[3, 13, 22, 24] Cultural and ethnic factors may affect the prevalence rate. This may be due to the influence of these factors on the likelihood of presentation to health care rather than real prevalence differences in the general population.[7, 21, 23] The typical presentation of physical complaints has varied throughout history and with the sociocultural environment of the patient.[32]

Aetiology

The aetiology of somatization disorder is unknown, but it is most likely multifactorial including biological, physiological, psychological, social, cultural, and iatrogenic factors. Different factors may have different importance at different times in the natural course of the illness. For instance, it may be a psychological trauma that precipitates the illness—but iatrogenic factors that maintain the illness. Whether the behavioural, cognitive, and other typical disturbances in somatization disorder and related disorders develop as a consequence of a basic biological defect in interaction with the patient's life experiences, or vice versa, or whether the mentioned disturbances and other predisposing factors have an independent impact is, as yet, unresolved.

Viewing functional somatic symptoms as a common reaction of human beings to stressors, like in the case of depressed mood and anxiety, results in the conclusion that different unspecific predisposing factors common for different mental disorders are involved.[33] Relatively specific predisposing factors are physical or sexual abuse and parental complaints of poor physical health and medically unexplained symptoms during the patient's childhood, whereas neither parental nor childhood well-defined physical illnesses seem to be predisposing factors.[34, 35]

The reported family transmission in somatization disorder may be due to sociocultural learning. However, there is some support for genetic transmission in somatization disorder, although twin studies have been inconclusive.[36]

Psychological theories

In the classical psychodynamic drive theory, medically unexplained physical symptoms are believed to develop as a reaction to the repression of unacceptable wishes or instinctual impulses and internal psychic conflicts.[37]

According to the theory of self-psychology, the anxiety connected with a threatening defragmentation or disintegration of the self is

the most profound form of anxiety that a person can experience.[37] In a defence against the feeling of emptiness, the individual becomes directed on the outside world and on physical stimuli. This process has been called 'stimulus entrapment' meaning that in somatization disorder, the individual becomes addicted to stimuli to his or her body.[38]

Individuals with alexithymia have a poorly developed language of emotions, and it has been suggested that instead they might respond with bodily symptoms. It is, however, unlikely that alexithymia has a specific aetiological role in somatization disorder.

The cognitive theory endorses the importance of the patients' misinterpretation of benign symptoms and normal physical sensations that they erroneously attribute to a physical disease.[39]

Biological factors

It is beyond doubt that an important biological component is involved, although the specific nature has still to be discovered. A neurophysiological dysfunction in the attention process has been demonstrated in somatization disorder, which may be explained by a reduced corticofugal inhibition in the diencephalon and the brainstem of afferent bodily stimuli, resulting in insufficient filtering of irrelevant bodily stimuli. A dysfunction of the secondary somatosensory area in the brain, a hypersensitivity of the limbic system towards bodily stimuli, or other dysfunctions may also be aetiologically involved.[40–42]

Other factors

In a few cases, a simple compensation claim may be of aetiological importance.

Physicians are primarily trained in a biomedical illness model and may have insufficient knowledge about diagnosing and managing somatization disorder and related disorders. Physicians may thus have a tendency to pursue organic possibilities and feel compelled to evaluate and treat all symptoms, and consequently a considerable iatrogenic reinforcement of physical symptoms is often involved.

Course and prognosis

Somatization disorder and related disorders have a spectrum of severity ranging from cases that may be difficult to delimit from normality to severely ill patients.[17]

In severe cases, the patients are chronically ill for most of their lives, but there may be periods of partial, but seldom full, remission. Some patients are able to work, others are severely disabled and are chair or bed bound, and their families have to provide virtually all aspects of physical care. Patients with somatization disorder and related disorders are often subjectively more functionally handicapped than patients who have a comparable, yet fully explained medical condition.[31]

Assessment

The purposes of the initial assessment are to (a) establish the diagnosis and rule out differential diagnoses, (b) examine which specific management or treatment strategies are possible and best for the patient, and (c) engage the patient in therapy.[15]

A scheme for the initial assessment by the psychiatric specialist is given in Table 5.2.3.3. For primary care physicians and in general hospitals this may be too comprehensive, and step one in the TERM model (Table 5.2.3.5) or another more simple model may be used.[43]

Table 5.2.3.3 Clinical assessment of patients with somatoform and related disorders including functional somatic syndromes

Before the meeting with the patient	
Review medical records and other relevant material	
At the examination	
Attitude towards the referral and the treatment	
Physical complaints	Chronology, intensity, provoking / relieving factors etc.
Triggering factors	Physical trauma or disease
	Psychosocial stressors
Current and previous emotional and behavioural complaints	
Social, funtional level, strain and coping	
The patient's illness belief and perception of symptoms	
Expectations to treatment and investigation	Physical Psychiatric
Past medical, surgical, and psychological history	
Dispositions	
Physical examination	
Paraclinical tests: Obtain focused diagnostic tests if not already done	

Before meeting the patient

Before meeting the patient it is important to review medical case notes and to gather information from other sources, e.g. the primary care physicians or the family, as patients with somatization disorder and related disorders may be inconsistent historians as a result of their complex medical history. The aim is to get an overview of the patient's medical history, the illness picture and complaints, examinations and diagnostic tests, and treatments and the outcome of these. It must also be assessed whether the patient is sufficiently examined for relevant differential diagnoses. The review may furthermore impart important information about psychological and social issues.

The examination

(a) Attitudes towards the referral and treatment

Patients with somatization disorder and related disorders may be sceptical about seeing a psychiatrist as they believe their problem to be of a physical nature and not psychiatric. This should be addressed directly by asking the patients what they have been told by the referring doctor, their reasons for coming to see the psychiatrist, and their feelings about it. It is important that such thoughts are brought to light to avoid misunderstandings and misconceptions, and to help the patients feel understood and in safe hands. In acknowledging the patients' fears of being stigmatized, it may be helpful to discuss negative public attitudes to psychiatry.

(b) Current physical complaints

The patients' physical complaints should be reviewed in detail for diagnostic purposes and to make the patients feel understood and taken seriously, which is a precondition for a good rapport between doctor and patient. Each symptom is explored to establish their characteristics, location, intensity, chronology and variation in intensity, onset, duration, and impact on daily life. Besides, provoking/relieving factors, previous treatments and the outcome of these are explored. It may be helpful together with the patient to write down the medical history on a time axis as patients may have difficulties with the chronology of their illness.

(c) Triggering factors

It is explored if onset of the illness is associated with physical trauma or diseases or with exposure to psychosocial stress or trauma.

(d) Current and previous emotional and behavioural complaints

As to a high comorbidity between somatization disorder and related disorders and other mental disorders, it is necessary to methodically clarify if the patients have a depression, anxiety disorder, or another mental illness.

(e) Illness beliefs and perception of symptoms

The patients' illness beliefs and perception of symptoms are of paramount importance for their illness and functioning as the illness behaviour has its origin in those beliefs and in illness attitudes. To change an inappropriate behaviour, it is necessary to identify such dysfunctional beliefs. Also, the attitudes and behaviour of the family may be a crucial factor in understanding the presentation and in planning the intervention.

(f) Expectations to treatments and investigations

The patients are questioned about wishes for and expectations to treatment and diagnostics tests and about what they believe may help them. Patients with somatization disorder and related disorders often have an unrealistic expectation to the impact of diagnostic tests and the effect of medical or surgical treatments. This may result in an intensive use of consultations. The patients must be helped to face the limits of what medicine can do and to acknowledge that continued medical consultations will be fruitless.

(g) Past medical, surgical, and psychological history

The patients' history is reviewed and related to the information gathered during the medical case note review. Furthermore, it is attempted to elucidate the patients' premorbid psyche and personality.

(h) Dispostions

Somatoform and related disorders often run in families and are associated with other mental disorders, and hence dispositions are explored.

(i) Physical examination

If not already made, a clinical examination ought to be carried out. Besides excluding organic possibilities, it also has a psychological purpose in making the patients feel that their physical complaints are taken seriously and that the psychiatrist is not exclusively focusing on the psychological part of the problem.

Routine laboratory test battery may be; complete blood count, electrolytes, blood urea nitrogen, creatinine, glucose, calcium, phosphate, liver function test, total protein, thyroid-stimulation hormone, erythrocyte sedimentation rate or CPR, urinalysis, and if indicated by symptoms or history, serological tests for Epstein–Barr virus, lyme disease, and immunological function test. Other tests may be relevant depending on the patient's illness picture.

Feedback of the results

Feedback of the psychiatric examination results to the referring doctor must be done in a way that is intelligible to a doctor who may not be psychologically minded. Statements like 'no formal psychiatric disorders are found', which unfortunately are frequent in consultation notes, may be somewhat useful when dealing with a patient referred for functional somatic symptoms. Such a statement just proves that the psychiatrist is unfamiliar with somatization disorder and related disorders.

The psychiatrist must be careful not to become involved in criticism of medical colleagues and in the divisions that these patients sometimes try to create between different therapists.

Treatment and management

Evidence-based treatment

Many different therapies have been used in patients with somatization disorder and related disorders like family therapy, physical therapy, biofeedback, relaxation therapies, hypnotherapy, psychodynamic psychotherapy, cognitive-behavioural therapy etc. The focus in the management varies a lot from (a) focus on the patients (organ-oriented approach or cognitive interpersonal approach, i.e. pattern of bodily and emotional symptoms over time, focus on dysfunction of central processing and context factors, interventions aimed at sensations, cognitions, affects, behaviours, and restoring overall function), (b) focus on the doctor (early recognition, communication skills, avoidance of iatrogenic harm), and (c) focus on context factors (doctor reimbursement system, patient compensation schemes, health care system, workplace characteristics, cultural belief).[44]

It must be concluded that the evidence for an effective treatment of patients with multiple functional symptoms is unsystematic. The use of multiple treatment methods and outcome measurement makes it difficult to compare studies. However, there seems to be substantial evidence that a specialized assessment with discharge letter, CBT, brief psychodynamic psychotherapy, and antidepressants have some effect on one or more outcome parameters.[44–47]

Treatment setting

Functional somatic symptoms and functional somatic syndromes are common in all medical settings. The patients often believe that they have a physical and not a psychological problem and will primarily seek medical and not psychological care. Hence, the management of somatization disorder and related disorders is not only an issue for psychiatrists but for all settings within the health care system.

The management must follow a stepped-care model, in which it is defined at which level of specialization each patient should be treated and, for each step, who is responsible for which parts of the treatment. For example, the mild and uncomplicated cases are treated by the primary care physician, modest to severe cases mainly by the primary care physician but in collaboration with a specialist, whereas the severe and complicated cases are managed in specialized care.

Besides the severity of the disorder, defining the steps in the model includes considerations about feasibility both as to available treatment resources and what is acceptable to the patient and the skills and knowledge of the primary care physician. Patients with a chronic somatization disorder may be well cared for by their primary care physician, provided the latter has the necessary skills and knowledge.

Inpatient care may be appropriate in a few cases, but patients with somatization disorder and related disorders are difficult to treat in ordinary psychiatric wards, where they are often met with considerable resistance from the psychiatric staff. Specialized inpatient units only exist in a few places in the world, and this treatment has not been documented.

Non-specialiszed treatment and management

(a) General hospital departments and non-psychiatric specialists

Because of the prevalence and the risk of iatrogenic harm, it is important that non-psychiatric specialists know about somatoform and related disorders and know how to identify and diagnose them. If the assessing physician exclusively focuses on symptoms that he finds relevant for his own specialty, there is a great risk of pursuing a wrong diagnosis. The fear of overlooking a definite physical disease, as an explanation of the physical symptoms, is deeply rooted in doctors, and this may, together with a poor knowledge about somatoform disorders, result in the doctor attempting to rule out even the rarest physical causes before a somatization disorder is even considered. However, there is little evidence that important medical diagnoses are missed more often in patients with somatoform disorders than in patients with other disorders.[48–50] Unnecessary procedures and diagnostic tests are not only unpleasant and potentially risky for the patient but may also delay or hinder sufficient treatment resulting in an aggravation of the disorder or perhaps chronicity. However, it must always be borne in mind that patients with a somatoform disorder may also have or acquire a concurrent physical disease. Instead of viewing somatization disorder and related disorders as a diagnosis of exclusion, all diagnostic possibilities ought to be included in the diagnostic consideration and the examination plan from the initial contact in the same way as when it is a question about two well-defined organic diseases. Diagnostic tests should be on medical indication and not on patient demand.

The primary role of the non-psychiatric specialists in the treatment of somatization disorder and related disorders is to:

- Exclude physical disease or trauma that can be treated medically or surgically
- In an empathic way, make it clear to the patient that he or she does not have the physical disease he or she fears, and that there is no indication of any other physical disease or defect that needs medical attention.
- That there is no medical indication for further diagnostic tests or examinations
- Coordinate the management with the primary care physician and other doctors that the patient may be in contact with.
- Consider a referral to a psychiatrist for examination or treatment
- In chronic cases, follow the advice given in Table 5.2.3.4.

Table 5.2.3.4 General advice on the non-specialized management of chronic somatization

Physical
1. Make a brief physical examination focusing on the organ system from which the patient has (new) complaints. - Look for signs of disease instead of symptoms. - Avoid tests and procedures unless indicated by signs of disease or a well-defined (new) clinical illness picture.
2. Reduce unnecessary drugs. Do not use on demand prescriptions and avoid dependence-forming medication.
Psychological
3. Make the diagnosis and inform the patient that the disorder is known and has a name.
4. Acknowledge the reality of the symptoms.
5. Be direct and honest with the patient about the areas you agree on, those you do not agree on, but be careful as not to make the patient feel ignorant or not respected.
6. Be stoical; do not expect rapid change or cure.
7. Reduce expectations of cure and accept the patient as being chronically ill. Aim at containment and (iatrogenic) damage limitation, i.e. use the management rather than treatment.
8. Perceive worsening of/or new symptoms as emotional communication rather than as a manifestation of a new disease.
9. Apply a specific therapeutic technique if you master it and consider referral to specialist treatment.
Psychopharmacological treatment
10. Consider treatment with psychoactive mediation (primarily antidepressant).
11. Choose non-habit forming medication and, if possible, choose medication that can be serum monitored.
12. Start with a smaller dosage than usual and increase slowly. Be stoical about side effects.
13. Take regular serum values to compliance issued and for validating complaints of adverse effects.
14. Treat any co-existing psychiatric disorders according to usual guidelines.
Administrative
15. Be proactive rather than reactive. Agree on a course with fixed, scheduled appointments with 2–6-weeks intervals and avoid consultations on patient demand (if needed, accept on demand a maximum of 1 phone appointment per week).
16. If the patient has a job, avoid giving him sick leave if at all possible.
17. Try to become the patient's only physician and minimize the patient's contact to other health care professionals, doctors on call, and alternative therapists.
18. Inform your colleagues of your management plans and develop contingency plans for when you are not accessible.
19. Inform the patient's nearest relative and try to co-opt a relative as a therapeutic ally.
20. If necessary, arrange support/supervision for yourself.
21. If necessary, motivate the patient to receive psychiatric treatment.

As previously mentioned, patients with somatization disorder are often resistant to psychiatric referral as they believe they have a physical and not a psychological problem. They may construe the referral as a sign that the physicians are not taking their symptoms seriously. It is important that the referring physicians avoid giving the message that the patients are not genuinely ill, that they trouble the doctors unnecessarily, or that they are 'mad'.[51] Instead, the

physician must try to meet the patients' wishes about knowing the cause of their illness, being taken seriously, getting explanations, information, advice, and reassurance.[52] A close liaison with medical colleagues guiding them in making a psychiatric referral in an acceptable way is important for engaging the patient in treatment.

Primary care

The TERM model (the extend reattribution model) is a simple cognitive-behavioural orientated treatment method to improve the primary care physicians' detection and management of patients presenting with medically unexplained symptoms. The method can effectively be taught to primary care physicians and will improve the outcome of their patients' treatment.[43] The TERM model is one of several different models that have been developed on the basis of the original reattribution model.[53,54]

The first stage of the TERM model (Table 5.2.3.5) is called 'understanding', as the important point, besides assessment of the patient, is that the patient feels understood and taken seriously by the doctor. The second stage is called 'the physician's expertise and acknowledgement of illness', in which the physician feeds back the results of his examination, but at the same time acknowledges the reality of the symptoms. The third stage is called 'reframing', in which a new model of understanding of the patient's problem is negotiated between patient and doctor. As a fourth stage, the model includes techniques for negotiating further treatment.[43]

Finally, the model includes principles for management of chronic, somatizing patients (Table 5.2.3.4). In chronic cases, damage limitation is a more realistic therapeutic goal than cure, and management is thus a more realistic aim than treatment.[51,55,56] The main aim is to stop the pathological cycle of interventions and consultations and the consequential somatic 'overtreatment' (i.e. treatment on obscure indication), and then, if possible, gradually to motivate the patient to accept specialized care if available in the area. Management according to these principles has shown to be effective in randomized controlled studies and should therefore always be implemented either solely or combined with one of the treatments described below.[57]

Table 5.2.3.5 The TERM model

1. **Understanding**
Take a full history of the symptoms
Explore emotional clues
Inquire directly about symptoms of anxiety and depression
Explore life events, stress, and other external factors
Explore functional level
Explore the patient's health beliefs
Explore the patient's expectations to treatment
Make a brief, focused physical examination
2. **The physician's expertise and acknowledgement of illness**
Provide feedback on the results of the physical examination
Acknowledge the reality of the symptoms
Make it clear that there is no (or that there is indeed) indication for further examination of nonpsychiatric treatment
3. **Negotiating a new model of understanding (reframing)**
Simple explanations
Physical symptoms are common reactions to, for example, stress and strain/nervousness
Depression lowers the threshold of pain
Muscular tension in anxiety and nervousness causes pain
Demonstrations
Practical (hyperventilation, muscular tension)
Establish the association between physical discomfort, emotional reactions, and life events
"Here and now" (nervous about consulting the physician)
Severe cases
Known phenomenon with a name: somatization
Basically the cause is unknown, but nothing indicates a hidden physical disease
Biological explanation: some are bodily sensitive than others, which explains their more intense symptoms
Individual symptom coping and reactions determine one's future well-being
4. **Negotiating further treatment**
Sum up agreements made during the consultation
Agree on specific objectives, contents, and form for the future course
Acute cases: no further appointments
Subacute cases: therapy sessions, regular scheduled appointments
Chronic: consider status consultation, regular scheduled appointments
Consider referral to psychiatrist, psychologist, or specialist service
5. **Chronic cases**
See Table 5.2.3.4

Specialized treatment and management

Cognitive-behavioural therapy is the best documented and most widely used therapy and is hence the focus in this section along with pharmacotherapy. The general principles of CBT are described elsewhere, so this section concentrates on the general techniques used in somatization disorder and related disorders, and how they differ from the CBT techniques applied in other disorders.

(a) Goal setting

Early in the therapy, the goals for the therapy are established. It is important to set up goals that are realistic in the light of the patient's illness and the framework of the therapy.

(b) Engagement and motivation

Treatment of somatization disorder and related disorders differs from the treatment of other mental disorders on several points, one being that it is very important to work systematically by engaging the patients in therapy. As the patients believe they have a medical condition, they may have very low belief in psychological treatment. It may be helpful to discuss the idea that all illnesses have an emotional component and that a psychological treatment focusing on the emotional component is often helpful in reducing suffering. The motivation and engagement should preferably be established during the assessment interview.

(c) Psychoeducation

The patient needs to be taught about somatoform and related disorders and about the body's normal reactions to stress and how stress may be expressed in physical symptoms. It is important that the patient learns about the possible biological and physiological basis in the CNS for the symptoms. This will make the reality of the symptoms clear to the patient and emphasize that the illness is not imaginary or made up. The education may be supported by written information. Parts of the information ought to be repeated during the therapy when it, in a relevant way, can be linked to the patient's personal experience.

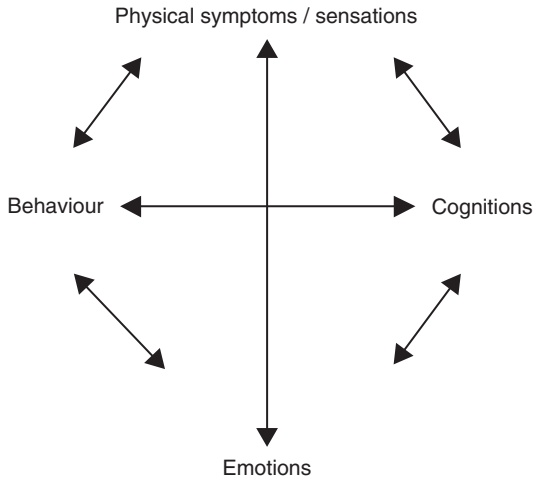


Fig. 5.2.3.1 Basic model in cognitive behavioural therapy.

(d) Physical symptoms and symptom attribution

Figure 5.2.3.1 illustrates the relationship between symptoms, cognitions, emotions, and behaviour in the cognitive-behavioural model. The first step in the therapy is to clarify the patient's dysfunctional automatic thoughts and basic beliefs about illness and symptoms, i.e. cause, consequences for future health and function, treatment, etc. Those thoughts are related to feelings and illness behaviour.

In the next step, the patient's disease model and symptom attributions are challenged. The therapist asks the patient to consider alternative possibilities, in which process even the most unlikely explanations are welcome. The therapist and the patient then explore which explanation is the most likely by investigating pros and cons for each possible explanation.

Finally, it is clarified how the patient would behave and react to a particular understanding or belief and which feelings this would produce.

(e) Behaviour and coping

Even if the patient does not want to or cannot work with his or her illness perception and symptom attribution, it would still be possible to work with the way the patient copes with illness. Illness behaviour and coping with symptoms and illness are scrutinized, and the behaviour is linked to the underlying thoughts and feelings if possible. When an example of a behaviour or coping strategy is elucidated, it is listed with arguments for and against a particular behaviour or coping strategy. A brainstorm on alternative possibilities is advisable, and pros and cons for each of these are also scrutinized. In a negotiation with the therapist, the patient chooses an alternative possibility, which is tested by the patient as homework, and at next session the effect of this is explored.

It is important to go slowly for the patient to experience success. The therapist is obliged to make sure that the patient sets up realistic goals with a good chance of success.

(f) Links between symptoms and stressors

The cause of somatization disorder and related disorder may be viewed as a combination of personal vulnerability and the stress and strain an individual is exposed to. The patients are often unaware of their patterns of reaction, but this can be established by careful registration of variations in symptom intensity and then relating them to what the patients are doing or thinking at that time. Based on variation in symptom intensity and stressors, potential important stressors in everyday life may be identified. This may provide a focus for intervention.

(g) Family and social network

The patient's health beliefs are often shared with the family and social network. Therefore it is necessary to create an alliance with the family to make sure they support the patient and not counteract the therapy. Family members are invited to a consultation and informed about the nature of the disorder and about the planned treatment. Misunderstandings and prejudgements about somatization disorder are eliminated.

(h) Treatment and help seeking behaviour and the physicians' handling

For patients with somatization disorder, social and family life is often centered on their illness, and an objective of the therapy is to reduce the importance of the illness and try to build up other interests and aspects of life. The patient's consulting behaviour may result in multiple fruitless diagnostic tests, referrals, and treatment attempts, which expose the patient to iatrogenic harm. Often, the patient grows tired of the doctors responding to their questions merely by referring or prescribing medication. In therapy, the patients are taught how to present their problems to the physicians in a way that prevents referrals and medication.[58] A patient may for instance tell the doctor that he or she is worried about some new symptoms, because he or she does not know whether the symptoms are just part of his or her somatoform disorder or something else, and that the doctor's expertise is required in order to clarify this.

Medication

Coexisting mood, anxiety, or other mental disorders are as effectively treated with psychotropic medication in patients with a somatoform disorder as in patients without.[59] Patients without coexisting mental disorders also seem to benefit from psychopharmacological treatment, antidepressants being the first choice.[44] In some cases, medication with peripheral action may be helpful as symptomatic treatment, for example in case of gastrointestinal symptoms like IBS.[44] Tricyclic antidepressants seem most effective, but due to side effects SSRI or SNRI ought to be the first choice. A useful strategy in the psychopharmacological treatment of somatizing patients is to start on a lower dosage than is usually recommended and to increase the dosage only gradually in order to avoid side effects. Antidepressants with the fewest interactions should be chosen as polypharmacy is common in these patients. The use of benzodiazepines and other dependence-producing drugs should be avoided. Stronger painkillers usually only produce a partial and temporary improvement, but they may result in misuse and should be avoided. Mild painkillers and tricyclic antidepressants are to be preferred.

In general, a degree of stoicism is required on the part of the clinician as patients may have varying symptom intensity and motivation and because patients may complain of side effects as a result of increased sensitivity to bodily sensations. Drugs that can be serum-monitored are preferred so that compliance and the likelihood of side effects can be assessed.

Possibilities for prevention

Patients with medically unexplained symptoms are often unpopular both with general psychiatrists and other doctors. The patients are seen as neither physically nor mentally ill, but simply as individuals complaining in order to avoid normal life responsibilities. Overcoming these negative attitudes is a matter of proper training of doctors and education of medical students.

Since an early diagnosis is important for preventing physical fixation, iatrogenic harm, and chronicity, doctors must be taught to view somatization disorder and related disorders with the same seriousness as well-defined physical diseases. The dominating dualistic way of thinking in medicine must be counteracted.

Further information

Mayou, R., Sharpe, M., and Carson, A.J. (eds.) (2004). *ABC of psychological medicine*. BMJ Books, London.

Woolfolk, R.L. and Allen, L.A. (2006). *Treating somatization. A cognitive-behavioral approach*. The Guilford Press.

Bass, C. (1990). *Somatization: physical symptoms & psychological illness* (1st edn). Blackwell Scientific Publications Oxford.

Mayou, R., Bass, C., and Sharpe, M. (1995). *Treatment of functional somatic symptoms.* Oxford University Press, Oxford.

References

1. Goldberg, D.P. and Bridges, K. (1988). Somatic presentations of psychiatric illness in primary care setting. *Journal of Psychosomatic Research*, **32**(2), 137–44.
2. Simon, G., Gater, R., Kisely, S., *et al.* (1996). Somatic symptoms of distress: an international primary care study. *Psychosomatic Medicine*, **58**(5), 481–8.
3. Fink, P., Toft, T., Hansen, M.S., *et al.* (2007). Symptoms and syndromes of bodily distress: an exploratory study of 978 internal medical, neurological, and primary care patients. *Psychosomatic Medicine*, **69**(1), 30–9.
4. Katon, W., Lin, E., Von Korff, M., *et al.* (1991). Somatization: a spectrum of severity. *American Journal of Psychiatry*, **148**(1), 34–40.
5. Perley, M.J. and Guze, S.B. (1962). Hysteria-the stability and usefulness of clinical criteria. A quantitative study based on a follow-up period of six to eight years in 39 patients. *The New England Journal of Medicine*, **266**, 421–6.
6. Murphy, M.R. (1990). Classification of the somatoform disorders. In *Somatization: physical symptoms & psychological illness* (ed. C. Bass), pp. 10–39. Blackwell Scientific Publications, Oxford.
7. Escobar, J.I., Rubio-Stipec, M., Canino, G., *et al.* (March 1989). Somatic symptom index (SSI): a new and abridged somatization construct. Prevalence and epidemiological correlates in two large community samples. *The Journal of Nervous and Mental Disease*, **177**(3), 140–6.
8. Kroenke, K., Spitzer, R.L., deGruy, F.V., *et al.* (1997). Multisomatoform disorder. An alternative to undifferentiated somatoform disorder for the somatizing patient in primary care. *Archives of General Psychiatry*, **54**(4), 352–8.
9. Kirmayer, L.J., Groleau, D., Looper, K.J., *et al.* (2004). Explaining medically unexplained symptoms. *Canadian Journal of Psychiatry*, **49**(10), 663–72.
10. Fink, P. (1992). Surgery and medical treatment in persistent somatizing patients. *Journal of Psychosomatic Research*, **36**, 439–47.
11. Fink, P. (1992). Physical complaints and symptoms of somatizing patients. *Journal of Psychosomatic Research*, **36**, 125–36.
12. Salmon, P., Peters, S., and Stanley, I. (1999). Patients' perceptions of medical explanations for somatisation disorders: qualitative analysis. *British Medical Journal*, **318**(7180), 372–6.
13. Stern, J., Murphy, M., and Bass, C. (1993). Personality disorders in patients with somatisation disorder. A controlled study. *British Journal of Psychiatry*, **163**, 785–9.
14. Fink, P. (1992). The use of hospitalizations by persistent somatizing patients. *Psychological Medicine*, **22**, 173–80.
15. House, A. (1995). The patient with medically unexplained symptoms: making the initial psychiatric contact. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 89–102. Oxford University Press, Oxford.
16. Yutzy, S.H., Cloninger, C.R., Guze, S.B., *et al.* (1995). DSM-IV field trial: testing a new proposal for somatization disorder. *The American Journal of Psychiatry*, **152**(1), 97–101.
17. Fink, P., Rosendal, M., and Olesen, F. (2005). Classification of somatization and functional somatic symptoms in primary care. *Australian and New Zealand Journal of Psychiatry*, **39**(9), 772–81.
18. Wiesmuller, G.A., Ebel, H., Hornberg, C., *et al.* (2003). Are syndromes in environmental medicine variants of somatoform disorders? *Medical Hypotheses*, **61**(4), 419–30.
19. Feldman, M.D. and Ford, C.V. (1994). *Patient or pretender. Inside the strange world of factitious disorders*. John Wiley & Sons, Inc, NY, USA.
20. Eisendrath, S.J. (1984). Factitious illness: A clarification. *Psychosomatics*, **25**, 110–17.
21. Üstün, T.B. and Sartorius, N. (1995). *Mental illness in general health care, an international study*. John Wiley & Sons, Chichester, UK.
22. Fink, P. (1995). Psychiatric illness in patients with persistent somatisation. *British Journal of Psychiatry*, **166**(1), 93–9.
23. Gureje, O., Simon, G.E., Ustun, T.B., *et al.* (1997). Somatization in cross-cultural perspective: a World Health Organization study in primary care. *The American Journal of Psychiatry*, **154**(7), 989–95.
24. Creed, F. and Barsky, A. (2004). A systematic review of the epidemiology of somatisation disorder and hypochondriasis. *Journal of Psychosomatic Research*, **56**(4), 391–408.
25. Fink, P., Steen, H.M., and Sondergaard, L. (2005). Somatoform disorders among first-time referrals to a neurology service. *Psychosomatics*, **46**(6), 540–8.
26. Fink, P., Hansen, M.S., and Oxhoj, M.L. (2004). The prevalence of somatoform disorders among internal medical inpatients. *Journal of Psychosomatic Research*, **56**(4), 413–18.
27. Toft, T., Fink, P., Oernboel, E., *et al.* (2005). Mental disorders in primary care: prevalence and co-morbidity among disorders. Results from the functional illness in primary care (FIP) study. *Psychological Medicine*, **35**(8), 1175–84.
28. De Waal, M.W., Arnold, I.A., Eekhof, J.A., *et al.* (2004). Somatoform disorders in general practice: prevalence, functional impairment and comorbidity with anxiety and depressive disorders. *British Journal of Psychiatry*, **184**, 470–6.
29. Kroenke, K. and Spitzer, R.L. (1998). Gender differences in the reporting of physical and somatoform symptoms. *Psychosomatic Medicine*, **60**(2), 150–5.
30. Barsky, A.J., Orav, E.J., and Bates, D.W. (2005). Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity. *Archives of General Psychiatry*, **62**(8), 903–10.
31. Hansen, M.S., Fink, P., Frydenberg, M., *et al.* (2002). Use of health services, mental illness, and self-rated disability and health in medical inpatients. *Psychosomatic Medicine*, **64**(4), 668–75.
32. Shorter, E. (1992). *From paralysis to fatigue. A history of psychosomatic illness in the modern era*. The Free Press, Macmillan Inc. New York.
33. Mayou, R., Bass, C., and Sharpe, M. (1995). Overview of epidemiology, classification, and aetiology. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 42–65. Oxford University Press, Oxford.
34. Hotopf, M., Carr, S., Mayou, R., *et al.* (1998). Why do children have chronic abdominal pain, and what happens to them when they grow

up? Population based cohort study. *British Medical Journal*, **316**(7139), 1196–200.
35. Leserman, J., Drossman, D.A., Li, Z., *et al.* (1996). Sexual and physical abuse history in gastroenterology practice: how types of abuse impact health status. *Psychosomatic Medicine*, **58**(1), 4–15.
36. Guze, S.B. (1993). Genetics of Briquet's syndrome and somatization disorder. A review of family, adoption, and twin studies. *Annals of Clinical Psychiatry*, **5**, 225–30.
37. Rodin, G.M. (1991). Somatization: a perspective from self psychology. *Journal of American Academy Psychoanalysis*, **19**, 367–84.
38. Meares, R. (1997). Stimulus entrapment: on a common basis of somatization. *Psychoanalytic Inquiry*, **17**(2), 223–34.
39. Sharpe, M. (1995). Cognitive behavioural therapies in the treatment of functional somatic symptoms. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 122–43. Oxford University Press, Oxford.
40. Miller, L. (1984). Neuropsychological concepts of somatoform disorders. *International Journal of Psychiatry in Medicine*, **14**(1), 31–46.
41. Rief, W. and Barsky, A.J. (2005). Psychobiological perspectives on somatoform disorders. *Psychoneuroendocrinology*, **30**, (10), 996–1002.
42. Mertz, H. (2002). Role of the brain and sensory pathways in gastrointestinal sensory disorders in humans. *Gut*, **51**(Suppl. 1), i29–i33.
43. Fink, P., Rosendal, M., and Toft,T. (2002). Assessment and treatment of functional disorders in general practice: the extended reattribution and management model—an advanced educational program for nonpsychiatric doctors. *Psychosomatics*, **43**(2), 93–131.
44. Henningsen, P., Zipfel, S., and Herzog, W. (2007). Management of functional somatic syndromes. *Lancet*, **369**(9565), 946–55.
45. O'Malley, P.G., Jackson, J.L., Santoro, J., *et al.* (1999). Antidepressant therapy for unexplained symptoms and symptom syndromes. *Journal of Family Practice*, **48**(12), 980–90.
46. Kroenke, K. (2007). Efficacy of treatment for somatoform disorders. A review of randomized controlled trials. *Psychosomatic Medicine*, in press.
47. Kroenke, K. and Swindle, R. (2000). Cognitive-behavioural therapy for somatization and symptom syndromes: a critical review of controlled clinical trials. *Psychothererapy and Psychosomatics*, **69**(4), 205–15.
48. Fink, P. (1997). *Persistent somatization*. Thesis; faculty of health sciences. University of Aarhus, Denmark.
49. Crimlisk, H.L., Bhatia, K., Cope, H., *et al.* (1998). Slater revisited: 6 year follow up study of patients with medically unexplained motor symptoms. *British Medical Journal*, **316**(7131), 582–6.
50. Stone, J., Wojcik, W., Durrance, D., *et al.* (2002). What should we say to patients with symptoms unexplained by disease? The 'number needed to offend'. *British Medical Journal*, **325**(7378), 1449–50.
51. Bass, C. (1990). Assessment and management of patients with functional somatic symptoms. In *Somatization: physical symptoms & psychological illness* (ed. C. Bass), pp. 40–72. Blackwell Scientific Publications, Oxford.
52. Price, J. and Leaver, L. (2002). ABC of psychological medicine: beginning treatment. *British Medical Journal*, **325**(7354), 33–5.
53. Gask, L. (1995). Management in primary care. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 391–409. Oxford University Press, Oxford.
54. Morriss, R., Gask, L., Ronalds, C., *et al.* (1998). Cost-effectiveness of a new treatment for somatized mental disorder taught to GPs. *Family Practice*, **15**(2), 119–25.
55. Smith, G.R. Jr. (1995). Treatment of patients with multiple symptoms. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp.175–87. Oxford University Press, Oxford.
56. Bass, C., Sharpe, M., and Mayou, R. (1995). The management of patients with functional somatic symptoms in the general hospital. In *Treatment of functional somatic symptoms* (eds. R. Mayou. C. Bass, and S. Sharpe), pp. 410–27. Oxford University Press, Oxford.
57. Smith, G.R. Jr., Rost, K., and Kashner, T.M. (1995). A trial of the effect of a standardized psychiatric consultation on health outcomes and costs in somatizing patients. *Archives of General Psychiatry*, **52**(3), 238–43.
58. Salmon, P., Humphris, G.M., Ring, A., *et al.* (2006). Why do primary care physicians propose medical care to patients with medically unexplained symptoms? A new method of sequence analysis to test theories of patient pressure. *Psychosomatic Medicine*, **68**(4), 570–7.
59. Katon, W. and Sullivan, M. (1995). Antidepressant treatment of functional somatic symptoms. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 103–21. Oxford University Press, Oxford.

5.2.4 Conversion and dissociation disorders

Christopher Bass

Introduction

Of all the disorders characterized by symptoms in the absence of disease, conversion disorders are perhaps the most difficult to explain. How, for example, can one explain functional blindness or a loss of function of both legs in the absence of conspicuous organic disease? The ancient Greeks recognized that if we suffer emotional disturbance as a result of some serious stress (such as personal injury or bereavement), this causes a change in the nervous system which leads in turn to symptoms in different parts of the body according to the underlying pathophysiology. Nineteenth century neurologists made significant advances when they identified specific ideas at the root of the symptoms. In the early nineteenth century Collie[1] also observed that the significance of, and attention to, a symptom or set of symptoms may depend more on what they mean (or their value) to the individual than on the biological underpinnings of the symptom itself.

Spence has recently argued that the problem in hysterical motor disorders is not the voluntary motor system *per se*: rather, *it is in the way that the motor system is utilized in the performance (or non-performance) of certain willed, chosen, actions.*[2] This model invokes a consciousness that acts upon the body and the world. By contrast, the psychodynamic ('conversion') model, which Freud introduced and which held sway for most of the twentieth century, invokes an unconscious mechanism 'acting' independently of consciousness, to interfere with voluntary movement. Spence has further argued that hysterical paralyses are maintained not by unconscious mechanisms, but by conscious processes. The maintenance of these symptoms requires the patient's attention, a characteristic of higher motor acts; the paralyses break down when the subject is distracted, consciousness is obtunded, or when it (the 'paralyses') is circumvented by reflexive motor routines. Hysterical paralyses, Spence avers, are quintessentially disorders of action (or inactions), which the patient disavows, when faced with some overwhelming situation, which threatens the identity of the self.[2]

One regrettable development of psychiatry's adoption of Freudian theory was the fracture in communication between the disciplines of psychiatry and neurology, which has only recently been restored by the sort of collaborative research currently being carried out by neurologists and psychiatrists.[3] In the last decade there have also been exciting advances in neuroimaging, which have stimulated research into the neurophysiology of hysteria, and these will be described later. This chapter will also emphasize contemporary approaches to management of these difficult clinical problems.

Problems with definition

There are a number of problems with the definition of the conversion disorders (CD). First, physical disorder must be excluded, but neurological co-morbidity is known to be high in patients with CD,[4] and distinguishing which symptoms are accounted for by organic disease and which are not can be difficult. Second, it is stated that a temporal association between a psychological stressor and the onset on the disorder should be identified, but in practice this is often impossible to establish and depends to a large extent on the skill of the interviewing doctor. Finally, by definition (according to the glossaries ICD-10 and DSM-IV)[5, 6]; the process should be unconsciously mediated, but it is difficult (some would say impossible) to distinguish between symptoms that are not consciously produced and those that are intentionally manufactured. The DSM-IV provides no criteria to distinguish conscious from unconscious intent, and many authors have argued that the criteria for whether the patients are consciously aware of producing these symptoms should be dropped from the diagnosis of CD.[7]

In clinical practice it is often difficult for a physician, faced with a patient in a hospital bed unable to use his or her legs despite normal tests and clinical findings, to differentiate between conversion disorder, factitious or fabricated disorder, or frank malingering. What the clinician is being asked to do is to determine whether or not the symptoms are being produced intentionally or not; and what the motives are. Table 5.2.4.1 attempts to provide a framework, but it highlights the shortcomings of psychiatric glossaries, which in turn expose the limitations of the medical model, which forces doctors to place patients in categories without taking into account the normal moral capacity of many individuals to exercise choice and determine (at least to some extent) their actions.[8] These medical conundrums have been explored in more detail in the chapter on factitious disorders and malingering (Chapter 5.2.9).

The role of volition

Central to recent debates about hysteria and conversion disorders is the extent to which a person's illness presentation is considered a product of free will and hence social deviance or the result of psychopathology and/or psychosocial influences beyond the volitional control of the subject.[9] The proposal that voluntary processes are involved in some way has a very long history: something prevents a specific voluntary behaviour from being executed through a 'negative', lack of movement (as in paralysis), or a 'positive', abnormality of movement (as in psychogenic tremor). If 'will' is regarded as a conscious capacity that humans possess to choose what to do or refrain from doing, then the problem in CD appears to be that the will fails to produce normal action.[10] Hence, the diagnostic importance is placed on the patient's veracity: if we believe him when he says that he cannot act normally we conclude that his will is impeded pathologically; if we do not believe him we conclude instead that his will is deployed to deceive us. This is the distinction required by the diagnostic systems.

Table 5.2.4.1 Relationships between conversion hysteria, factious disorder, and malingering

	Subject insight		Target of deception		Perceived outcome	Motivation/ reason
	Aware	Unaware	Conscious self	Other		
Hysterical conversion		+	+		Sick and disabled role	Care/ dependency
Factious disorder	+			+	Sick and disabled role	Care/ dependency
Malingering	+			+	Sick and disabled role	Personal benefit, e.g. financial, avoiding prison

(Reproduced from Halligan, P. Bass, C. and Oakley, D. Wilful deception as illness behaviour. In *Malingering and illness deception* (eds. P. Halligan, C. Bass, and D. Oakley), pp. 3–28. Copyright 2003, with permission from Oxford University Press).

Conversion and dissociation

The word *conversion*is conventionally applied to somatic symptoms whereas if the symptom is psychological (e.g. a loss of memory or an external hallucination) rather than bodily (e.g. a loss of power) it is regarded as dissociative. *Dissociation* has attracted considerable recent interest, and it has been argued that the available evidence is more consistent with a model that identifies at least two distinct categories of dissociative phenomena—'detachment' and 'compartmentalization'—that have different definitions, mechanisms, and treatment implications.[11] These have been referred to as Type 1 (compartmentalization) and Type 2 (detachment), respectively (see Table 5.2.4.2).

Compartmentalization phenomena are characterized by impairment in the ability to control processes or actions that would usually be amenable to such control and which are otherwise functioning normally. This category encompasses unexplained neurological symptoms (including dissociative amnesia) and benign phenomena such as those produced by hypnotic suggestion. By contrast, *detachment phenomena* are characterized by an altered state of consciousness associated with a sense of separation from the self, the body, or the world. Depersonalization, derealization and out-of-body experiences constitute archetypal examples of detachment in this account. Evidence suggests that these phenomena are generated by a common pathophysiological mechanism involving the top-down inhibition of limbic emotional processing by frontal brain systems. Although these two types of dissociation are typically conflated, evidence suggests that different pathological mechanisms may be operating in each case.

Table 5.2.4.2 Classification of two types of pathological dissociation

Type 1 dissociation (compartmentalisation)	Type 2 dissociation (detachment)
Conversion disorders	Depersonalization/derealization
Dissociative amnesia	Peri-traumatic dissociation
Dissociative fugue	Out of body experiences
Dissociative identity disorder	Autoscopy (?)

(Reproduced with permission, from R. Brown (2002))

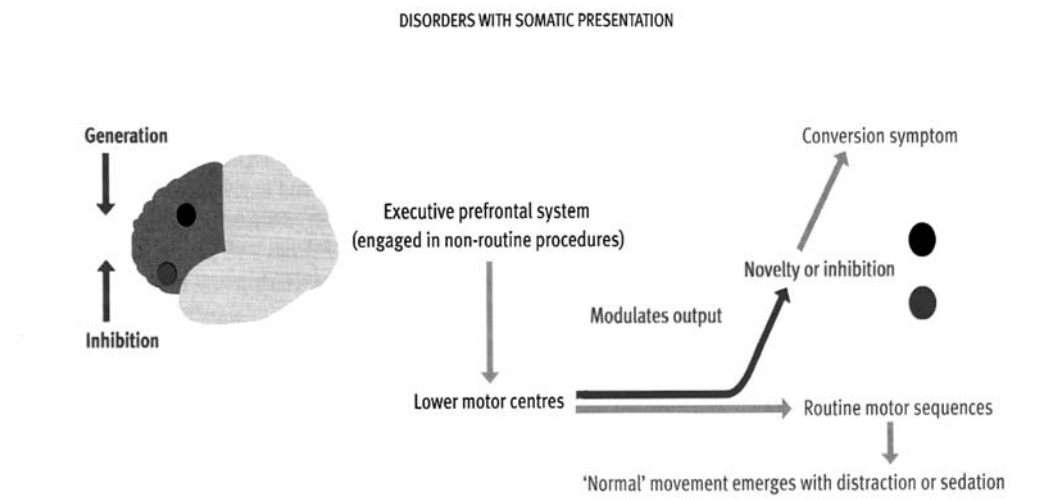
Support for the compartmentalization model comes from psychophysiological research, which suggests that psychogenic illness is associated with a deficit in attentional, conscious processing and the preservation of preattentive, preconscious processes. According to Brown[12] there is very little difference between 'negative' symptoms such as sensory loss, paralysis, etc. and 'positive' symptoms such as tremor, dystonia, etc. in terms of basic underlying mechanisms. By this view, all symptoms result from a loss of normal high-level attentional control over low-level processing systems; in this sense, all symptoms can be thought of as involving a form of compartmentalization.

Pathophysiology

There has been considerable progress in cognitive neuroscience and functional imaging over the last decade, which has provided a conceptual and empirically based platform for developing a neuroscience of not only hysterical symptoms but also free will.[13,14]

Recent functional neuroimaging data suggest that neural circuits linking volition, movement, and perception are disrupted in CD.[13] There are many studies examining the role of specific prefrontal regions in action generation (particularly the dorsolateral prefrontal and supplementary motor areas) and action suppression (especially the orbitofrontal cortices). These 'higher' executive centres supervene only when a change of behaviour is required: inappropriate behaviour must be suppressed or difficult procedures attended to, as when concentration is necessary. Hence, if the problem in hysteria is one of the will, and of abnormality emerging only when subjects attend to their actions, then this suggests the hypothesis that the prefrontal cortex is pivotal to the conversion process (see Fig. 5.2.4.1).

Further evidence that the prefrontal cortices play a key role in the control of action comes from a study of a woman with a left-sided conversion disorder affecting her leg. Marshall *et al.*[15] demonstrated that her attempt to move her paralysed leg was associated with increased activation of orbitofrontal (inhibitory) prefrontal regions, in the absence of motor cortical activity. They argued for an inhibition of motor behaviour by higher centres. Spence and colleagues[16] demonstrated that in three men with conversion symptoms affecting their upper limbs, hypokinetic movement was associated with reduced activation of dorsolateral (action-generation) areas of prefrontal cortex. Moreover, these areas of hypoactivity differed from those exhibited by four healthy men who were asked to feign the same motor impairments (see Fig. 5.2.4.2). It is possible that the application of functional neuroimaging techniques might allow clinicians to distinguish conversion from feigning on objective, empirical grounds.



The executive supervenes on lower motor centres when there is need for novel action generation (black circle) or suppression of inappropriate action (red circle), each implicating specific prefrontal regions: the dorsolateral and orbitofrontal cortices, respectively. Under conditions where the executive may be hypothesized to be disengaged (by distraction or sedation) normal movements emerge (as would be expected of routine actions). Conversion movements seem to require attention, and hence the engagement of the executive. There are 2 mechanisms by which conversion might emerge: failure to generate new actions, consequent upon dorsolateral hypofunction (black circle) or suppression of ongoing motor action, secondary to orbitofrontal activation (red circle). Other mechanisms may also operate. Acknowledgement: Mrs Jean Woodhead.

Fig. 5.2.4.1 Schematic diagram illustrating the role of prefrontal executive in modulating lower motor systems, and its hypothesized involvement in conversion disorder. (Reproduced with permission from Spence, S. (2006). Hysteria: a new look. In *Psychiatry*, **5**(2), pp. 56–60, Elsevier Ltd.)

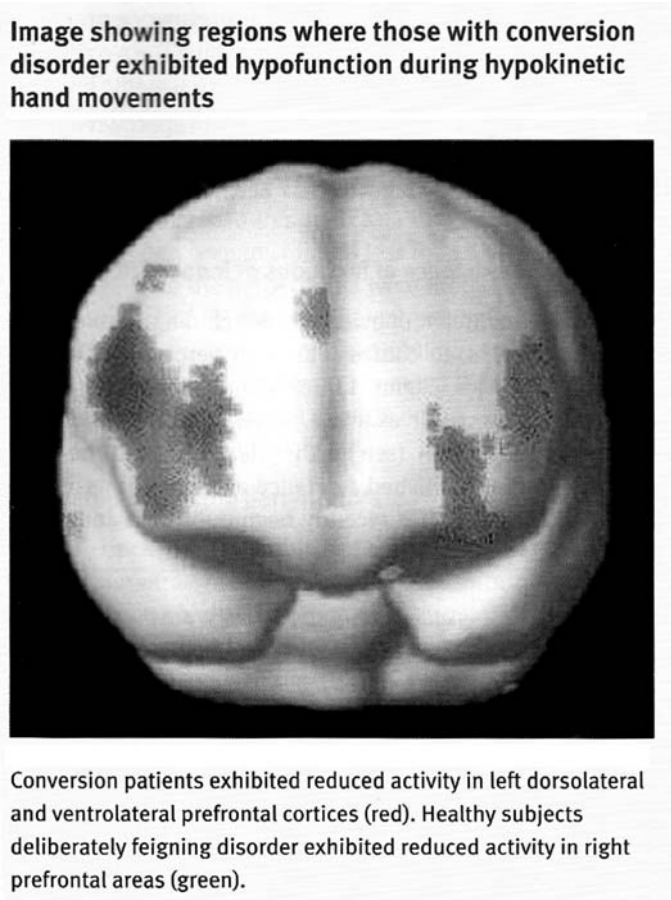


Fig. 5.2.4.2 Image showing regions where those with conversion disorder exhibited hypofunction during hypokinetic hand movements.
(Reproduced with permission (sought) from Spence, S. (2006). Hysteria: a new look. In *Psychiatry*, **5**(2), pp. 56–60, Elsevier Ltd.

In another recent case report using fMRI a patient with right-sided paralysis was asked to recall traumatic memories using a standard life event schedule: cued recall of the event was associated with regional brain activities characteristic of emotional arousal, including the amygdala and right inferior frontal lobe. Such recall was also associated with reduced motor activity in the area corresponding to the subjective paralysed limb.[17] This case study provides neuroimaging evidence for a connection between traumatic events and ongoing neurological symptoms (see Problems with definition, above).

Epidemiology

It is ironic that these research advances have occurred at a time when social historians have confidently asserted that hysteria has disappeared from clinical practice:

The most consequential development in the history of hysteria in the last century was the rapid decline in the medically recorded incidence of the disorder . . . . . .Hysteria—considered variously as a term, a theory, and a behaviour—is vanishing.[18]

It is extraordinary that this was written in 2001, at a time when symptoms considered 'functional, psychogenic, medically unexplained or hysterical' account for up to a third of new referrals to neurology outpatient departments.[3] and up to 9 per cent of admissions to a UK neurology in patient ward.[19] Akagi and House[20] concluded that the lowest prevalence figures suggested a rate of about 50 per 100 000 for cases of CD known to health services at any one time, with perhaps twice that number affected over a 1-year period. These figures suggest that hysteria is as common as other disabling conditions such as multiple sclerosis and schizophrenia. Furthermore, the burden of disability associated with chronic hysteria is far higher than a typical practising psychiatrist might expect, or than is reflected in standard textbooks of psychiatry or clinical neurology.[21]

It is regrettable therefore that it receives none of the resources or media attention that these disorders attract.

Clinical features

Conversion disorder: motor symptoms

The most typical motor symptoms are paralyses, functional weakness, gait disturbances, fits resembling epilepsy, and abnormal movements.

In the last decade diagnostic procedures have improved and the availability of non-invasive, accurate imaging has drastically reduced the rates of undetected organic pathology in patients with diagnoses of hysteria. Indeed, several recent studies have reported rates of misdiagnosis of between 0 and 4 per cent in regional and tertiary neurological centres,[22] which suggests that a diagnosis of CD can be made relatively confidently and accurately. In the following section the process of diagnosis will be briefly outlined through the history, examination, and investigation.

(a) The history

The onset, temporal sequence, and character of the presenting complaint may not be typical of a neurological disorder, and a number of other features may emerge, especially after an interview with a family member or a review of the hospital and general practitioner notes. If the patient is admitted to a general hospital bed the psychiatrist should routinely telephone the patient's primary care doctor and request a recent print out of his/her medical records (with consent). These often reveal key information about life events and/or antecedent illnesses, investigations, etc.

(i) Age of onset and sex

The average age of onset is the mid-30s, and patients with functional paralysis are less likely to be female than patients with pseudo seizures.

(ii) Mode of onset/recent life events or difficulties

An increased number of life events in the year preceding symptom onset have been recorded in small controlled studies of unexplained motor symptoms[23,24]. When patients are interviewed carefully, some report symptoms of panic just before the onset of, for example, functional weakness.[25] Judicious questions about sensations of sweating, dizziness, and difficulty breathing may reveal these somatic symptoms of anxiety, which may also be reported before the onset of sensory symptoms (see below). There is also an important literature describing unilateral somatic symptoms (which may present with sudden onset of functional weakness in a limb) following hyperventilation/panic.[26,27] These patients may present acutely in the A and E department of a general hospital, where they

may be admitted to the hospital stroke unit, or be sent to the general hospital as an emergency with an accompanying letter from the patient's GP describing the patient as 'off legs-please see and investigate.' The following clinical vignette is typical:

A 40-year-old woman was admitted as an emergency to the Stroke unit of the general hospital after collapsing at home. She had been involved in a dispute with her employers for weeks, and on the morning of the referral had an argument with her mother which made her very upset, and during the course of this she became distressed and developed paraesthesiae down the left side of her body, slurred speech, and collapsed, losing consciousness for 30 s. An ambulance was called and she was admitted to the A and E department, where she was noted to be hyperventilating and agitated, and to have weakness of the left leg, unintelligible speech, and claimed not to be able to see. All neurological investigations were normal, and after 12 days in hospital she made a gradual recovery, but was left with functional weakness in the left leg (requiring a wheelchair), and her speech was intermittently 'child-like'. History revealed a similar episode 3 years previously, a recent 2-year history of treatment for irritable bowel syndrome, and considerable work and domestic stress.

She was followed up by the liaison service and community mental health team, but despite the efforts of a community physiotherapist, psychologist, and nursing support her limb weakness continued and she remained disabled; 1 year after admission she was in receipt of disability benefits.

(iii) Previous unexplained symptoms

Evidence is accumulating that the more unexplained symptoms the patient has, the more likely the primary symptom is to be unexplained by disease.[28] In a recent study of patients with medically unexplained motor symptoms, additional unexplained symptoms including paraesthesia (65 per cent), pseudo epileptic seizures (23 per cent), and memory impairment (20 per cent) were reported.[4] It is often useful therefore, when CD is being considered as a diagnosis, to obtain a print out of the patient's past history from the primary care doctor (having obtained the patient's consent). This may reveal repeated presentations to different specialists as well as a history of repeated surgical procedures, particularly without clear evidence of pathology. It is also worth noting that patients with a diagnosis of somatization disorder (what used to be referred to as Briquet's syndrome) have high rates of conversion disorders, which punctuate their illness careers, often after a life event or physical injury/procedure.[29]

(iv) Psychiatric co-morbidity

Rates of depression (38 to 50 per cent) and anxiety (10 to 16 per cent) have been identified in a number of studies. In a recent small prospective controlled study there was a fourfold increase in depression in comparison with matched control with similar organic disability.[30]

(v) Neurological disease and other physical factors

A diagnosis of functional paralysis can be made in a patient who already has some paralysis from another cause, for example, the 'disproportionate disability' in a patient with multiple sclerosis. In a recent study of patients with unexplained motor symptoms 42 per cent had a co-morbid neurological disease and half of these had a peripheral origin.[4] Epilepsy is thought to coexist in a significant percentage of patients with non-epileptic attacks.[31]

It is also important for the physician or surgeon to be aware of the diverse ways in which conversion symptoms can present in the general hospital. In the last decade the author has seen many patients with conversion disorders after surgical procedures, investigations, and operations such as hysterectomy, minor injuries in the workplace, and after (often trivial) road traffic accidents. Numerous case reports implicating accidents, minor surgical procedures, and general anaesthetics as initiating factors have been described, for patients with both non-epileptic seizures and functional neurological syndromes.[32–34] These presentations are more likely to be seen in a medico legal setting, where the symptoms are shaped by the prospect of financial gain.

(vi) Secondary gain/litigation

This is a complex issue but impending litigation has been described in a number of studies of patients with unexplained motor symptoms and tremor.[4,35] One group of patients who may develop abnormal movements are those with reflex sympathetic dystrophy (RSD) which has been renamed Complex Regional Pain Syndrome Type I (CRPS I). It has been reported that patients with CRPS I with abnormal movements typically exhibit pseudo neurological (non-organic) signs, and in some cases malingering has been documented by secret surveillance.[36] The authors concluded that abnormal movements in CRPS I are a key clinical feature that differentiates CRPS I from CRPS II. Psychiatrists will often be asked to express opinions on patients with chronic painful extremities (often labelled as 'CRPS') in which abnormal movements have developed, especially in a medico legal setting. Pearce[37] has remarked that CRPS is best construed as a reaction to injury, or to excessive, often iatrogenic, immobilization after injury; but should not be seen as an independent disease. He asserts that the diagnosis of CRPS groups together ill-defined symptoms under a convenient, but medically untestable label, and that patients, lawyers, and support groups commonly deny psychogenesis, with the sadly mistaken notion that this implies a bogus or spurious cause.

(vii) Laterality of the symptoms

The idea that left-sided symptoms are more common than right has a long history but a recent systematic review found no evidence to support this view.[38]

(viii) History from relative/informant

There is a considerable amount of evidence to suggest that the observations and attitudes of carers may be important in the perpetuation of medically unexplained symptoms, especially motor conversion symptoms. For example, Davison *et al.*[39] found carers to be ill-informed and dissatisfied with the advice they had received from doctors about their relatives' diagnosis and disabilities. The education of carers and relatives is essential and will be dealt with in the section on management.

(b) The examination and diagnostic discrepancies

There is often a discrepancy between the patient's concept of the symptoms and the physician's knowledge of the anatomy and physiology. The way in which a patient moves or undresses may indicate a global affection that is incompatible with a specific nerve lesion or with a hemiplegia.

When considering any sign of functional weakness it is important to remember the following caveats:

1 Any sign that depends on inconsistency does not distinguish 'hysterical' from 'malingered' weakness.

2 The presence of a positive sign of functional weakness does not exclude the possibility that the patient also has an organic disease as well.

3 All physical signs, whether for organic or non-organic signs, have a limited reliability and inter-rater reliability.[22]

Another myth wedded to the concept of hysteria (I have already referred to the mistaken but common belief that it has 'disappeared' and that symptoms are more common on the left) is that the patients exhibit 'la belle indifference' or are inappropriately under concerned about their symptoms. In a recent systematic review Stone *et al.*[40] found that the median frequency of la belle indifference was 21 per cent (range 0–54 per cent) in 356 patients with conversion symptoms and 29 per cent (range 0–60 per cent) in 157 patients with organic disease. Indifference to symptoms is more likely to be noted in patients with factitious disorder (see Chapter 5.2.9).

Give way weakness is often used as a diagnostic test of hysterical paralyses, but it is unreliable. Unilateral functional weakness of a leg, if severe, tends to produce a characteristic gait in which the leg is dragged behind the body as a single unit, like a sack of potatoes. The hip is either held in external or internal rotation so that the foot points inwards or outwards. The most impressive quantitative discrimination to date between hysterical and neurological weakness is reported in a study of Hoover's sign—the involuntary extension of hysterically paralysed leg when the 'good leg' is flexing against resistance. Ziv and colleagues[41] demonstrated a clear difference in the pattern of response between neurological and psychogenic patient groups. It should be borne in mind however that the patient may have both a functional and organic disorder.[4]

Individual symptoms

(a) Paralyses

Paralyses may affect one or more limbs, or one side of the face. They may be flaccid or occur with contractures. In hysterical spasm both arm and leg are contracted on the same side of the body, the hand is closed tightly, the knee is flexed, and perhaps the leg and the foot are drawn up. Paralysis with contractures is one of the most extreme examples of disability caused by hysterical illness.

Hysterical paraplegia has been described,[42] and both spinal and orthopaedic surgeons, as well as rehabilitation specialists and neurologists, should be alert to the development of this disorder in their patients.[43] These patients have the potential to use considerable health care resources.[21]

(b) Abnormal movements

Psychogenic movement disorders are thought to account for 1 in 30 of all patients attending a movement disorder clinic[35] and have been the subject of a recent book.[44] During the last two decades a number of case series of patients with psychogenic dystonia have been reported.[45,46] Clinical features that suggest a psychogenic movement disorder are shown in Table 5.2.4.3.

In a recent systematic study of patients with fixed dystonia Schrag *et al.*[47] found that 37 per cent fulfilled criteria for psychogenic dystonia and 29 per cent criteria for *somatization disorder*, which is characterized by chronic, multiple, persistent, medically unexplained symptoms. Despite the fact that many patients fulfilled strict criteria for a somatoform disorder/psychogenic dystonia, in a proportion of patients the diagnosis remained uncertain, and whether the disorder was primarily neurological or psychiatric remains an open question. These patients require the services of a multi-disciplinary team.

Table 5.2.4.3 Features that suggest a psychogenic movement disorder

Abrupt onset
- Inconsistent movements (changing characteristics over time)
- Incongruous movements (movements do not fit with recognized patterns or with normal physiological patterns)
- Presence of additional types of abnormal movements that are not consistent with the basic abnormal movement pattern or are not congruous with a known movement disorder, particularly:
  - Rhythmical shaking
  - Bizarre gait
  - Deliberate slowness in carrying out the requested voluntary movement
  - Bursts of verbal gibberish
  - Excessive startle (bizarre movements in response to sudden, unexpected noise or threatening movement)
- Entrainment of the psychogenic tremor to the rate of the requested rapid successive movement the patient is asked to perform
- Demonstrating exhaustion and fatigue
- Spontaneous remissions
- Movements disappear with distraction
- Response to placebo, suggestion, or psychotherapy
- Presence as a paroxysmal disorder
- Dystonia beginning as a fixed posture

(Adapted with permission, from S. Fahn (1995))

The most common form of psychogenic movement disorder however is *psychogenic tremor*.[48] Almost 75 per cent of presenting patients are female and preceding events include work-related injuries and other accidents. A positive entrainment test (see Table 5.2.4.3), absence of finger tremor, and slowness of voluntary movements are suggestive of psychogenic origin. One-third has co-morbid somatoform disorders and one-fifth is involved in litigation or compensation. Prognosis is relatively poor if the condition has persisted for over 1 year, and in the long-term 80–90 per cent of patients continue to have abnormal movements.

(c) Seizures (psychogenic non-epileptic seizures or PNES)

It is estimated that more than 25 per cent of patients receiving a diagnosis of refractory epilepsy in a chronic epilepsy clinic do not have epilepsy.[49] Although the population incidence of PNES may be only 4 per cent that of epilepsy, PNES comprises a large share of the workload of neurologists and emergency and general physicians. Unlike patients with epilepsy however, those with PNES often do not have designated nurses or health care workers assigned to help with the management of this potentially disabling disorder.

PNES can be distinguished from epileptic seizures: PNES generally occur in the presence of an audience or when one is close at hand. They may be precipitated by stress, but more often seem to occur in response to the social setting. The fall to the ground is not usually abrupt, and movements may follow the fall with clutching, but the characteristic regular tonic–clonic sequence of epilepsy is not found. Tongue biting and incontinence of urine are rare in

hysterical fits, the corneal reflexes are preserved and the plantars are flexor, unless previously abnormal. Firm handling and pressure on the supra orbital nerves to the point of pain may arouse the patient. PNES occur most often among epileptic patients or among others who have seen epileptic fits. A few epileptic patients learn how to induce ictal discharges and can produce extra fits. Although rarely available during a fit, the EEG is generally abnormal in epilepsy and normal during hysterical fits.[31]

If PNES is not diagnosed and managed early, significant iatrogenic harm may occur. The outcome is not always favourable in these patients: in one recent study carried out at a mean of 11.9 years after manifestation and 4.1 years after diagnosis of PNES, 71 per cent of patients continued to have seizures and 56 per cent were dependent on social security. Outcome was better in patients with greater educational attainments, younger onset and diagnosis, attacks with less dramatic features, and fewer additional medically unexplained complaints.[50]

It has recently been reported that patients with PNES have a consistently different psychosocial profile from patients with motor conversion symptoms. In a prospective study of consecutive neurological inpatients with either motor conversion or pseudo seizures of recent onset, patients with PNES were younger, more likely to have both an emotionally unstable personality disorder and a lower perception of parental care, to report incest, and to have reported more life events in the 12 months before symptom onset than patients with motor conversion symptoms.[24] Recently a helpful fact sheet has been produced to help patients with PNES, which explains the nature of the disorder and approaches to management.[51] Although cognitive–behavioural therapy has been shown to be helpful in an open trial of patients with PNES, these findings need replication in a controlled setting.[52]

Sensory symptoms

(a) Sensory disturbance

The clinical detection and localization of sensory dysfunction is probably one of the least reliable areas of the neurological examination. Sensory loss may involve half the entire body from top to toe or from right to left. It may affect the whole of a limb, and characteristically has a glove or stocking distribution on the arms or legs, or both. The sensory loss generally fails to fit in with known anatomical boundaries but conforms more with the patient's concept of physiology and anatomy. Thus hysterical sensory loss is likely to stop sharply at the midline, while non-hysterical sensory change will only approach the midline since at this point segmental nerves overlap by one or two centimetres on each side.

Unfortunately these classical signs are often unreliable. 'Psychogenic' features on sensory examination and diminished vibration sense over the affected part of the forehead have been found in over half of patients with neurological disorders.[53, 54] Rolak also found that 'midline splitting' of sensory function was not helpful in determining whether there was an underlying neurological disorder.[54] These clinical findings should clearly be interpreted with circumspection.

Toth has recently described 34 patients with the 'hemisensory' syndrome, in which patients present with hemisensory disturbance and intermittent blurring of vision in the ipsilateral eye (asthenopia) and sometimes ipsilateral hearing problems as well.[55] Hemisensory symptoms are increasingly recognized in patients with chronic pain and in patients with reflex sympathetic dystrophy.

(b) Visual disturbances

Ophthalmologists have estimated that psychogenic visual disorders account for up to 5 per cent of their practice.[56] Simple observation of visually guided behaviour will sometimes reveal telling inconsistencies, particularly in the case of severe apparent visual loss. A number of reliable optometric techniques are available to support bedside tests and the diagnosis of psychogenic visual loss, field disturbance, or gaze abnormality (for more details see Stone *et al.*[23]). Disabling hysterical blindness presents more difficulties. Evoked potential studies will help to demonstrate intact visual pathways.

The disability associated with chronic conversion disorders

This topic is under-researched, but it is worth noting that, as the prognosis of CD is often poor, (not infrequently because patients are not diagnosed promptly, and even after diagnosis there are no resources to treat the patient). In the experience of the author this clinical conundrum is not improving: with the increase in provision of psychiatric services to the community and those with 'serious mental illness', patients with CD, even if they are profoundly disabled, often do not receive the appropriate treatment. Much of the disability is iatrogenic, and these patients will, not infrequently, be referred to the psychiatric service after having become confined to a wheelchair and/or in receipt of long-term disability benefits.[21,39] By this time the patients are usually entrenched in the sick role and it is very difficult to change the status quo.

Prognosis

The aetiological implications accruing from recent follow-up studies[4, 57] suggest that a short history and young age are held to be predictors of good outcome, while the presence of a personality disorder, chronicity of symptoms, receipt of disability benefits, and involvement with litigation predict poor recovery. As regards social circumstances, a change in marital status, good family functioning, and the elimination of a stressor has been shown to have a positive effect on outcome.[58] There is little chance of improvement once the symptoms have become chronic and enduring.

Patients with chronic motor symptoms, e.g. unilateral functional weakness, as well as those with sensory symptoms, appear to do particularly poorly. In particular, patients with unexplained motor symptoms who are referred to tertiary care centres continue to do very poorly following discharge. Despite the stability of the diagnosis, a pattern of multiple hospital referrals continues for many of these patients once they have been discharged from the tertiary care centre. Interviews of patients conducted on an average of 6 years after their original admission to a tertiary care centre revealed that many continued to be referred to neurologists and other specialists, but that subsequent psychiatric referral was rare.[59] Many changed their primary care doctor after discharge from hospital and a disproportionate number of re-referrals were made by primary care doctors who had known their patients for less than 6 months. Psychological attribution of symptoms was rare, and many patients felt dissatisfied with the treatment they had received. Many were exposed to unnecessary iatrogenic harm. These consistent findings of very poor outcome following discharge from neurological outpatient and inpatient services in patients with both

unexplained motor disorders as well as PNES suggest that without appropriate treatment the prognosis is poor.

Management

It is remarkable that a disorder as common as schizophrenia and multiple sclerosis should have attracted so little research interest or treatment resources. One reason for this is that there have been no randomized controlled treatment studies of CD, and so at the time of writing there is no good evidence about the best intervention for conversion disorder. There has been considerable interest in this topic however, which has been the subject of a recent Cochrane review.[60] All of the studies in this review were of poor methodological quality. On the credit side there is evidence that interest in CD is increasing and attracting more research funding.

Resources

Before any discussion of treatment it is important to consider the resources available to the neurologist to manage these patients. It is anomalous that, unlike disorders such as MS and schizophrenia, which have a similar prevalence, there are no designated resources for these patients. Some neurologists may have no access whatever to mental health resources, whereas others may have close collaborative links with either clinical psychology or psychiatry services. There is no doubt that the successful management of these patients requires the co-operation of a number of clinical specialties, including psychologists, nurses, physiotherapists, and occupational therapists (OTs). Some patients may be so disturbed or disabled (or both) that they may require inpatient admission to a specialized unit with access to both, mental health care and medical nurses, as well as physiotherapists and OTs. In the opinion of this writer every neurology service should have access to a specialist liaison psychiatry service.[61]

Management strategies for the neurologist

First, the diagnosis has to be established by a neurologist after relevant organic disease has been excluded. Second, the neurologist has to not only explain to the patient that there is no serious underlying organic disease but also provide an explanation for the symptoms that is comprehensible to the patient.

It is worth noting at this stage that patients prefer the term 'functional' rather than 'hysterical', when their unexplained weakness, fits, etc. are being referred to.[62] It is also important to avoid verbal landmines—for example using the phrase 'not sinister' instead of 'not serious'; or 'not structural' instead of 'not physical'. In patients who are generally hostile to psychological explanations it is best to use the word 'functional' instead of 'psychological'.

An example of an explanation to a patient with functional weakness and sensory disturbance may be something like: *you have what we call functional weakness. This is a common medical problem. Your nervous system is not damaged—we can see that from the examination and scans, etc. This is why why when you try and send the messages to your limbs they do not move properly. Similarly this is why the sensations from your body are not being felt properly. The most important thing about this condition is that because your nervous system is not damaged, the problem is potentially reversible. All the parts of the nervous system are there but are just not working properly, so that when you try to move your leg it doesn't do it as well as it should. Sometimes stress can cause these symptoms, which are often accompanied by worry and low mood but these are not the cause of the problem. Stress is a common problem and can lead to headache and abdominal pain as well as what we call functional weakness.*

This explanation can be supplemented by giving the patient a **fact sheet** containing information about functional weakness, which contains information about how to become involved with rehabilitation[51] (Fig. 5.2.4.3).

Further management

Ideally the neurologist and psychiatrist should interview the patient together at the bedside, but this is not always possible. At the very least however close collaboration between the two is essential before the patient is reviewed by the psychiatrist and a formulation proposed (and any potential for iatrogenic illness or diagnostic confusion eliminated).

Traditional behavioural approaches to treatment are based on the premise that the symptoms reported by the patient are interpreted as physical but are amenable to recovery. Treatment aims to bring about a gradual increase in function through a combination of physical and occupational therapies. The patient receives rewards and praise for improvement of function, and withdrawal of reinforcement for continuing signs of disability. Avoiding direct confrontation of psychological problems and providing 'face-saving' techniques are also regarded as key components.[63] More recently the approach to patients has moved from a predominantly medical one, to one in which psychological and sociocultural aspects are equally important, and the need for organized specialist rehabilitation services involving a multi-disciplinary team is recognized as essential.

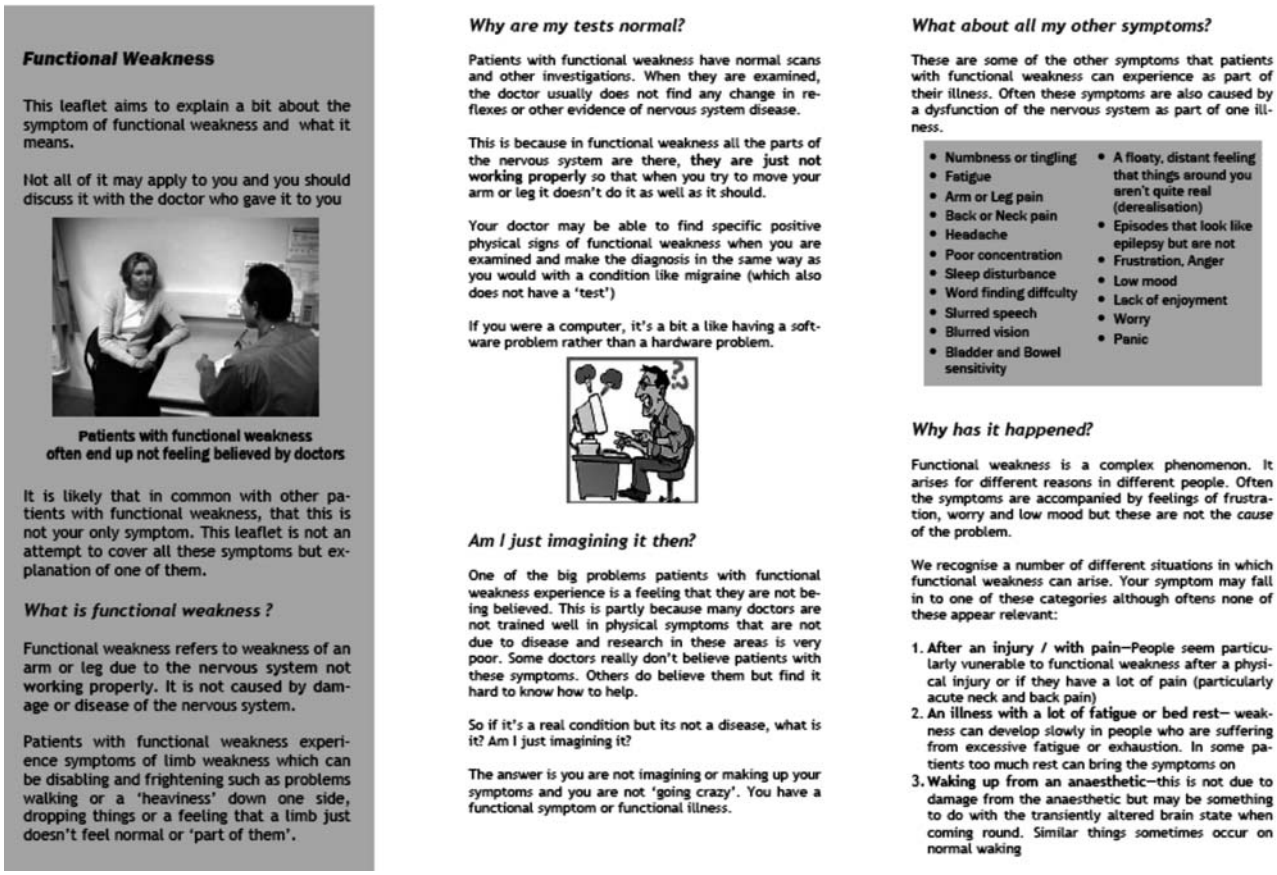
What is the evidence?

With one or two recent exceptions,[64] there are no large, randomized controlled studies of treatment in patients with CDs. Neither is there any good evidence to support the use of one specific intervention, e.g. biofeedback, hypnosis, psychotherapy. Although repeated case series have documented the effectiveness of multidisciplinary inpatient behavioural treatment, there is little controlled research.

In the absence of good experimental evidence a possible framework for future research has been developed which is based on published evidence and described in the WHO ICIDH.[65] This is particularly useful for patients in whom there is a disability that is out of proportion to known disease and signs. The model provides opportunities for intervention, and is well suited to the kind of multi-disciplinary approach that is likely to be successful in these patients.

The model emphasizes that whatever the primary cause of an illness, many factors (both individual and systemic) will have an influence on its manifestations. *For example: a vignette here.*

A 20-year-old woman was referred to the liaison service with functional paraplegia of 12 months duration. Onset was temporally related to a back strain caused by lifting a chair. She was confined to a wheelchair and lived with her parents and 14-year-old brother in an adapted house (specially adapted chair and stair rails) and was in receipt of benefits. A home visit was carried out and the patient denied any current problems or recent life events, although she described a long history of medically unexplained symptoms, multiple food allergies, and previous treatment in an adolescent unit for chronic



Functional Weakness

This leaflet aims to explain a bit about the symptom of functional weakness and what it means.

Not all of it may apply to you and you should discuss it with the doctor who gave it to you

Patients with functional weakness often end up not feeling believed by doctors

It is likely that in common with other patients with functional weakness, that this is not your only symptom. This leaflet is not an attempt to cover all these symptoms but explanation of one of them.

What is functional weakness ?

Functional weakness refers to weakness of an arm or leg due to the nervous system not working properly. It is not caused by damage or disease of the nervous system.

Patients with functional weakness experience symptoms of limb weakness which can be disabling and frightening such as problems walking or a 'heaviness' down one side, dropping things or a feeling that a limb just doesn't feel normal or 'part of them'.

Why are my tests normal?

Patients with functional weakness have normal scans and other investigations. When they are examined, the doctor usually does not find any change in reflexes or other evidence of nervous system disease.

This is because in functional weakness all the parts of the nervous system are there, **they are just not working properly** so that when you try to move your arm or leg it doesn't do it as well as it should.

Your doctor may be able to find specific positive physical signs of functional weakness when you are examined and make the diagnosis in the same way as you would with a condition like migraine (which also does not have a 'test')

If you were a computer, it's a bit a like having a software problem rather than a hardware problem.

Am I just imagining it then?

One of the big problems patients with functional weakness experience is a feeling that they are not being believed. This is partly because many doctors are not trained well in physical symptoms that are not due to disease and research in these areas is very poor. Some doctors really don't believe patients with these symptoms. Others do believe them but find it hard to know how to help.

So if it's a real condition but its not a disease, what is it? Am I just imagining it?

The answer is you are not imagining or making up your symptoms and you are not 'going crazy'. You have a functional symptom or functional illness.

What about all my other symptoms?

These are some of the other symptoms that patients with functional weakness can experience as part of their illness. Often these symptoms are also caused by a dysfunction of the nervous system as part of one illness.

- Numbness or tingling
- Fatigue
- Arm or Leg pain
- Back or Neck pain
- Headache
- Poor concentration
- Sleep disturbance
- Word finding diffculty
- Slurred speech
- Blurred vision
- Bladder and Bowel sensitivity
- A floaty, distant feeling that things around you aren't quite real (derealisation)
- Episodes that look like epilepsy but are not
- Frustration, Anger
- Low mood
- Lack of enjoyment
- Worry
- Panic

Why has it happened?

Functional weakness is a complex phenomenon. It arises for different reasons in different people. Often the symptoms are accompanied by feelings of frustration, worry and low mood but these are not the *cause* of the problem.

We recognise a number of different situations in which functional weakness can arise. Your symptom may fall in to one of these categories although oftens none of these appear relevant:

1. **After an injury / with pain**—People seem particularly vunerable to functional weakness after a physical injury or if they have a lot of pain (particularly acute neck and back pain)
2. **An illness with a lot of fatigue or bed rest**— weakness can develop slowly in people who are suffering from excessive fatigue or exhaustion. In some patients too much rest can bring the symptoms on
3. **Waking up from an anaesthetic**—this is not due to damage from the anaesthetic but may be something to do with the transiently altered brain state when coming round. Similar things sometimes occur on normal waking

Fig. 5.2.4.3 Fact sheet for patient with functional weakness (Reproduced from Stone, J. Carson, A. and Sharpe, M. (2005), Functional symptoms in neurology: management. *Neurology in Practice,* **71**(Suppl. 1), i13–i21. Copyright 2005, BMJ Publishing Group Ltd.)

fatigue syndrome. She was seeing a chiropractor for her symptoms, and had been told that she had nerve damage (not confirmed). Her brother was off school with chronic fatigue and her mother had a long history of emotional problems. At interview, the family were polite but could not identify any link between recent events and her disabling symptoms. Discussion with the GP did not reveal any other relevant information and a follow-up was arranged 1 month later. Before this appointment the patient telephoned the psychiatrist to say that she had not disclosed certain facts at the initial meeting, and revealed that she had confronted her physically abusive step father and asked him to leave the house to live with his mistress (which he duly did). At follow-up there was a great deal of expressed relief, and the patient agreed with our formulation that, to some extent, she had developed weakness in the legs and become wheelchair bound in order to avoid being hit by her step father. She agreed to a brief admission to a rehabilitation unit, was encouraged to mobilize gradually and had regular sessions with a clinical psychologist. She responded well to treatment, commenced a college course, and learned to drive a car. At 3 year follow-up she was symptom free and in gainful employment. [the first vignette in this chapter].

Psychological treatments

Because patients with conversion disorders share features in common with patients with other medically unexplained syndromes, treatments that have been used in these latter disorders may have potential. Most of the evidence-based treatments in this field involve cognitive–behavioural therapy (CBT,[66] or interpersonal therapy (IPT)). These usually have to be undertaken by trained clinical psychologists or other clinicians. However, increasing numbers of specialist nurses are being trained to deliver these treatments, so they should become more widely available.

CBT is concerned mainly with helping the patients overcome identified problems and ascertain specified goals. It discourages 'maintaining factors' such as repeated body self-checking and excessive bedrest, and challenges patients' negative or false beliefs about symptoms. Chalder has described specific CBT based treatment for patients with conversion disorders.[67]

Hypnosis and intravenous sedation

In an inpatient trial of hypnosis both patients with CD and controls improved equally and no extra effect from hypnosis was found.[68] Others have found the use of intravenous sedatives, particularly propofol, helpful in persuading some patients with whom a good relationship has been established, that they can eventually make a recovery.[51]

Pharmacological treatments

There is evidence from randomized controlled trials (RCTs) and systematic reviews that antidepressants (both tricyclics and selective

serotonin reuptake inhibitors (SSRIs)) can be useful in the treatment of patients with medically unexplained symptoms (such as poor sleep and pain), whether or not depression is present.[69]

Further information

Halligan, P., Bass, C., and Marshall, J. (eds.) (2001). *Contemporary approaches to the study of hysteria. Clinical and theoretical perspectives*. Oxford University Press, Oxford.

Hallett, M., Fahn, S., Jancovic, J., *et al.* (eds.) (2005). *Psychogenic movement disorders: psychobiology and treatment of a functional disorder*. Lippincott, Williams and Wilkins.

Trimble, M. (2004). *Somatoform disorders: a medicolegal guide*. Cambridge University Press, Cambridge.

References

1. Collie, J. (1913). *Malingering and feigned sickness*. Edward Arnold, London.
2. Spence, S. (1999). Hysterical paralyses as disorders of action. *Cognitive Neuropsychiatry*, **4**, 203–26.
3. Carson, A., Ringbauer, B., Stone, J., *et al.* (2000). Do medically unexplained symptoms matter? A prospective cohort study of 300 new referrals to neurology outpatient clinics. *Journal of Neurology, Neurosurgery, and Psychiatry*, **68**, 207–11.
4. Crimlisk, H., Bhatia, K., Cope, H., *et al.* (1998). Slater revisited: 6 year follow up study of patients with medically unexplained motor symptoms. *British Medical Journal*, **316**, 582–6.
5. ICD-10 Classification of Mental and Behavioural Disorders (1992). WHO, Geneva.
6. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn.). American Psychiatric Association, Washington, DC.
7. Shapiro, A. and Teasell, R.W. (2004). Behavioural interventions in the rehabilitation of acute v chronic non-organic (conversion/factitious) motor disorders. *British Journal of Psychiatry*, **185**, 140–6.
8. Bass, C. and Halligan, P. (2007). Illness related deception: social or psychiatric problem? *Journal of the Royal Society of Medicine*, **100**, 81–4.
9. Halligan, P., Bass, C., and Oakley, D. (2003). Wilful deception as illness behaviour. In *Malingering and illness deception* (eds. P. Halligan, C. Bass, and D. Oakley), pp. 3–28. Oxford University Press, Oxford.
10. Spence, S., Hunter, M., and Harpin, G. (2002). Neuroscience and the will. *Current Opinion in Psychiatry*, **15**, 519–26.
11. Holmes, E., Brown, R., Mansell, W., *et al.* (2005). Are there two qualitatively distinct forms of dissociation? A review and some clinical implications. *Clinical Psychology Review*, **25**, 1–25.
12. Brown, R.J. (2004). Psychological mechanisms of medically unexplained symptoms: an integrative conceptual model. *Psychological Bulletin*, **130**, 793–812.
13. Broome, M. (2004). A neuroscience of hysteria? *Current Opinion in Psychiatry*, **17**, 465–9.
14. Ghaffar, O., Staines, W., and Feinstein, A. (2006). Unexplained neurologic symptoms: an fMRI study of sensory conversion disorder. *Neurology*, **67**, 2036–8.
15. Marshall, J., Halligan, P., Gink, G., *et al.* (1997). The functional anatomy of a hysterical paralysis. *Cognition*, **64**, B1–8.
16. Spence, S., Crimlisk, H., Cope, H., *et al.* (2000). Discrete neurophysiological correlates in prefrontal cortex during hysterical and feigned disorder of movement. *Lancet*, **355**, 1243–4.
17. Kanaan, R.A., Wessely, S., and David, A. (2007). Imaging repressed memories in motor conversion disorder. *Psychosomatic Medicine*, **69**, 202–5.
18. Micale, M. (2001). Hysteria. In *The Oxford companion to the body* (eds. C. Blakemore and S. Jennett), pp. 382–4. Oxford University Press, Oxford.
19. Parry, A., Murray, B., Hart, Y., *et al.* (2006). Audit of resource use in patients with non-organic disorders admitted to a UK neurology unit. *Journal of Neurology, Neurosurgery and Psychiatry*, **77**, 1200–1.
20. Akagi, H. and House, A. (2002). The clinical epidemiology of hysteria: vanishingly rare, or just vanishing? *Psychological Medicine*, **32**, 191–4.
21. Allanson, J., Wade, D., and Bass, C. (2002). Characteristics of patients with persistent severe disability and medically unexplained neurological symptoms: a pilot study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **73**, 307–9.
22. Stone, J. and Zeman, A. (2001). Hysterical conversion-a view from clinical neurology. In *Contemporary approaches to the science of hysteria* (eds. P. Halligan, C. Bass, and J. C Marshall) pp. 102–25, Oxford University Press, Oxford.
23. Stone, J., Sharpe, M., and Binzer, M. (2004). Motor conversion symptoms and pseudoseizures: a comparison of clinical characteristics. *Psychosomatics*, **45**, 492–9.
24. Binzer, M., Stone, J., and Sharpe, M. (2004). Recent onset pseudoseizures—clues to aetiology. *Seizure*, **13**, 146–55.
25. O'Sullivan, G., Harvey, I., Bass, C., *et al.* (1992). Psychophysiological investigations of patients with unilateral symptoms in the hyperventilation syndrome. *British Journal of Psychiatry*, **160**, 664–7.
26. Tavel, M. (1964). Hyperventilation syndrome with unilateral somatic symptoms. *Journal of the American Medical Association*, **187**, 301–3.
27. Blau, N., Wiles, M., and Solomon, F. (1989). Unilateral somatic symptoms due to hyperventilation. *British Medical Journal*, **286**, 1108.
28. Wessely, S., Nimnuan, C., and Sharpe, M. (1999). Functional somatic syndromes: one or many? *Lancet*, **354**, 36–9.
29. Bhui Hotopf, M. (1997). Somatisation disorder. *British Journal of Hospital Medicine*, **58**, 145–9.
30. Binzer, M., Andersen, P., and Kullgren, G. (1997). Clinical characteristics of patients with motor disability due to conversion disorder: a prospective control group study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **63**, 83–8.
31. Reuber, M. and Elger, C. (2003). Psychogenic non epileptic seizures: review and update. *Epilepsy & Behaviour*, **4**, 205–16.
32. Letonoff, E.J., Williams, T.R., and Sidhu, K. (2002). Hysterical paralysis: a report of three cases and a review of the literature. *Spine*, **27**, E441–5.
33. Lichter, I., Goldstein, L., Toone, B., *et al.* (2004). Nonepileptic seizures following general anaesthetic: a report of 5 cases. *Epilepsy & Behaviour*, **5**, 1005–13.
34. Reuber, M., Howlett, S., Khan, A., *et al.* (2007). Non-epileptic seizures and other functional neurological symptoms: predisposing, precipitating, and perpetuating factors. *Psychosomatics*, **48**, 230–8.
35. Factor, S., Podskalny, R., and Molho, E. (1995). Psychogenic movement disorders: frequency, clinical profile and characteristics. *Journal of Neurology, Neurosurgery and Psychiatry*, **59**, 406–12.
36. Verdugo, R. and Ochoa, J. (2000). Abnormal movements in complex regional pain syndrome: assessment of their nature. *Muscle Nerve*, **23**, 198–205.
37. Pearce, J. (2005). Chronic regional pain and chronic pain syndromes. *Spinal Cord*, **43**, 263–8.
38. Stone, J., Sharpe, M., Carson, A., *et al.* (2002). Are functional motor and sensory symptoms really more frequent on the left? A systematic review. *Journal of Neurology, Neurosurgery, and Psychiatry*, **73**, 578–81.
39. Davison, P., Sharpe, M., Wade, D., *et al.* (1999). "Wheelchair" patients with non organic disease: a psychological enquiry. *Journal of Psychosomatic Research*, **47**, 93–103.
40. Stone, J., Smyth, R., Carson, A., *et al.* (2006). La belle indifference in conversion symptoms and hysteria: systematic review. *British Journal of Psychiatry*, **188**, 204–9.
41. Ziv, I., Djaldetti, R., and Zoldan, Y. (1998). Diagnosis of "non-organic" limb paresis by a novel objective motor assessment: the quantitative Hoover's test. *Journal of Neurology*, **245**, 797–802.
42. Baker, J. and Silver, J. (1987). Hysterical paraplegia. *Journal of Neurology, Neurosurgery and Psychiatry*, **50**, 375–82.

43. Heruti, R., Reznik, J., Adunski, A., *et al.* (2002). Conversion motor paralysis disorder: analysis of 34 consecutive referral. *Spinal Cord*, **430**, 335–40.
44. Hallett, M., Fahn, S., Jankovic, J., *et al.* (eds.) (2005). *Psychogenic movement disorders: psychobiology and treatment of a functional disorder*. Lippincott, Williams and Wilkins, Philadelphia.
45. Fahn, S. and Williams, D. (1988). Psychogenic dystonia. In *Advances of Neurology*, Vol. 50: *Dystonia 2* (eds. S. Fahn, *et al.*), pp. 431–55. Raven Press, New York.
46. Lang, A.E. (1995). Psychogenic dystonia: a review of 18 cases. *Canadian Journal of Neurological Sciences*, **22**, 136–43.
47. Schrag, A., Trimble, M., Quinn, N., *et al.* (2004). The syndrome of fixed dystonia: an evaluation of 103 patients. *Brain*, **127**, 2360–72.
48. Bhatia, K. and Schneider, S. (2007). Psychogenic tremor and related disorders. *Journal of Neurology*, Apr 9; [Epub ahead of print].
49. Smith, D., Defalla, B., and Chadwick, D. (1999). The misdiagnosis of epilepsy and the management of refractory epilepsy in a specialist clinic. *Quarterly Journal of Medicine*, **92**, 15–23.
50. Reuber, M., Pukrop, R., Bauer, J., *et al.* (2003). Outcome in psychogenic nonepileptic seizures: 1 to 10 year follow up in 164 patients. *Annals of Neurology*, **53**, 305–11.
51. Stone, J., Carson, A., and Sharpe, M. (2005). Functional symptoms in neurology: management. *Neurology in Practice*, **71**(Suppl. 1), i13–i21. http://www.jnnp.com.
52. Goldstein, L., Mellors, J., and Toone, B. (2004). An evaluation of cognitive behavioral therapy as a treatment for dissociative seizures: a pilot study. *Cognitive Behavioural Neurology*, **17**, 41–9.
53. Gould, R., Miller, B., Goldberg, M., *et al.* (1986). The validity of hysterical signs and symptoms. *Journal of Nervous and Mental Diseases*, **174**, 593–7.
54. Rolak, L. (1988). Psychogenic sensory loss. *Journal of Nervous and Mental Disease*, **176**, 686–7.
55. Toth, C. (2003). Hemisensory syndrome is associated with a low diagnostic yield and a nearly uniform benign prognosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, **74**, 1113–6.
56. Kathol, R., Cox, T., Corbett, J., *et al.* (1983). Functional visual loss: I. A true psychiatric disorder? *Psychological Medicine*, **13**, 307–14.
57. Stone, J., Sharpe, M., Rothwell, P., *et al.* (2003). The 12-year prognosis of unilateral functional weakness and sensory disturbance. *Journal of Neurology, Neurosurgery, and Psychiatry*, **74**, 591–6.
58. Ron, M. (2001). The prognosis of hysteria/somatisation disorder. In *Contemporary approaches to the study of hysteria* (eds. P. Halligan, C. Bass, and J. Marshall), pp. 271–83, Oxford University Press, Oxford.
59. Crimlisk, H., Bhatia, K., Cope, H., *et al.* (2000). Patterns of referral in patients with medically unexplained motor symptoms. *Journal of Psychosomatic Research*, **49**, 217–9.
60. Ruddy, R. and House, A. (19 October 2005). Psychosocial interventions for conversion disorder. *Cochrane Database Systematic Review*, (4): CD005331.
61. Gotz, M. and House, A. (1998). Prognosis of symptoms that are medically unexplained. *British Medical Journal*, **317**, 536.
62. Stone, J., Wojcik, W., and Durrance, D., (2002). What should we say to patients with symptoms unexplained by disease? The number needed to offend. *British Medical Journal*, **325**, 1449–50.
63. Teasell, R. and Shapiro, A. (1993). Rehabilitation of chronic motor conversion disorder. *Critical Review of Physical and Rehabilitation Medicine*, **5**, 1–13.
64. Moene, F., Spinhoven P, *et al.* (2003). A randomized controlled clinical trial of a hypnosis-based treatment for patients with conversion disorder, motor type. *International Journal of Clinical and Experimental Hypnosis*, **51**, 29–50.
65. WHO. (2001). International classification of functioning, disability and health. World Health Organization, Geneva. http://www3.who.int/icficftemplate.cfm
66. Kroenke, K. and Swindle, R. (2001). Cognitive behavioural therapy for somatisation and symptom syndromes: a critical review of controlled clinical trials. *Psychotherapy and Psychosomatics*, **69**, 205–15.
67 Chalder, T. (2001). Cognitive behavioural therapy as a treatment for conversion disorders. In *Contemporary approaches to the study of hysteria: clinical and theoretical perspectives* (eds. P. Halligan, C. Bass, and J. Marshall), pp. 298–311. Oxford University Press, Oxford.
68. Moene, F., Spinhoven, P., Hoogduin, K., *et al.* (2002). A randomised controlled clinical trial on the additional effect of hypnosis in a comprehensive treatment programme for inpatients with conversion disorder of the motor type. *Psychotherapy and Psychosomatics*, **71**, 66–76.
69. Jackson, J., O'Malley, P., and Kroenke, K. (2006). Antidepressants and cognitive-behavioral therapy for symptom syndromes. *CNS Spectrum*, **11**, 212–22.

5.2.5 Hypochondriasis (health anxiety)

Russell Noyes Jr.

Introduction

Hypochondriasis is a preoccupation with the fear that one has, or may develop, serious disease despite evidence to the contrary. So defined, the disorder affects between 2 and 7 per cent of patients attending general medical clinics and is a cause of physical dysfunction and disability.[1] It is also a reason for increased health care utilization and dissatisfaction with care received. To their physicians, patients with this disorder are an enigma and a source of frustration.

Unfortunately, relatively little is known about hypochondriasis. Primary care physicians have had little interest and psychiatrists see few patients with the condition. It is a pejorative label that, even if entertained, is rarely communicated. And, even if communicated, the diagnosis would not, until very recently, have led to effective treatment.

History

Hypochondria was used by Hippocrates to refer to a region below the cartilage of the ribs. In the second century, Galen linked it to organs in this area as well as humours and animal spirits. The symptom picture was ill-defined and only gradually took on the characteristics recognized today. From earliest times the disorder was associated with melancholia, a temperamental disturbance caused by an excess of black bile. Burton (1621) described hypochondriacal melancholy in terms of vague physical symptoms, disturbances of mood, and fears. In the seventeenth century, Sydenham viewed hypochondria in men as the counterpart of hysteria in women, but the first modern description was published in 1799 by Sims.

By the eighteenth century, hypochondria became part of a fashionable disturbance that Cheyne attributed to the English way of life and environment. However, as notions of aetiology began to shift under the influence of Cartesian dualism, hypochondria was increasingly seen as a weakness and moral failing. Falret (1822) was

Table 5.2.5.1 Essential and associated features of hypochondriasis

Essential features
Fear of disease
Disease conviction
Bodily preoccupation
Somatic symptoms
Reassurance-seeking
Associated features
Fear of aging and death
Overvaluation of health
Low self-esteem
Sense of vulnerability to illness

perhaps the first to identify it as a mental disorder, one of the neuroses. Freud viewed hypochondria as an 'actual neurosis', having a physiological basis and not amenable to psychoanalysis. However, present-day descriptions began with Gillespie,[2] who in 1928 defined hypochondriasis as 'a mental preoccupation with a real or supposititious physical or mental disorder'.

Conceptualizations

Authors disagree about how hypochondriasis should be conceptualized. Some look upon it as a personality trait; its early onset and long-term stability in many patients fit this conception. Others view it as a dimension of psychopathology. They see illness worry as a continuum with hypochondriasis falling on the severe end. For those who take a categorical approach, the issue of whether hypochondriasis is primary or secondary remains unsettled. High rates of comorbidity create doubt about its independent status. Based on existing evidence, some question whether hypochondriasis can be regarded as a discrete psychiatric disorder.[3]

Clinical picture

Essential features

The essential characteristics of hypochondriasis are shown in Table 5.2.5.1. These include fear of serious disease, the consequences of which may include pain, suffering, disability, and death. Such fears take the form of alarming thoughts and images of specific diseases. They also include conviction or belief that the feared disease is already present. This belief is overvalued meaning that it is strongly held despite lack of evidence; it is not delusional.

Bodily preoccupation is perhaps the most important feature.[4] This takes the form of intense interest in, and attention to, what is happening in the body. The focus is upon somatic symptoms which tend to be multiple and diffuse. Attention is also directed to bodily sensations, bodily functions, and minor abnormalities as well as related concerns such as diet, exercise, and environmental exposures. The activities and conversation of patients are dominated by medical concerns. As a consequence of their self-absorption, interest in other people and pursuits is withdrawn.

Reassurance-seeking is the main behavioural feature. Patients repeatedly check their bodies for signs of serious disease. They check their pulse, look for lumps, examine themselves in the mirror, etc. In addition, they search medical sources for the meaning of their symptoms. Such patients also ask friends, family, and medical professionals for reassurance. Their search may lead to excessive utilization of health services.

Associated features

Associated characteristics include fears of aging and death, which appear to be an integral part of hypochondriasis. Overvaluation of health and appearance is another related feature. Hypochondriacal patients may become preoccupied with eating natural foods, achieving physical fitness, and living a healthy lifestyle, activities that reflect their idealized conception of good health.

Patients with hypochondriasis feel unworthy and unlovable.[4] As a consequence of their low self-esteem they have negative expectations of others including medical professionals. In addition, they have a sense of vulnerability to illness.[5] These characteristics have to do with fundamental aspects of the self that the hypochondriacal patient views as deficient.

Subtypes

Hypochondriacal patients are heterogeneous and subtypes may exist. Separate dimensions of disease phobia and disease conviction have consistently been identified; in some patients fears are prominent and in others conviction dominates the picture. Others may resemble patients with obsessive-compulsive disorder or personality disorders of one kind or another.

Classification

Criteria

Hypochondriasis initially appeared in DSM-II as one of the neuroses. In DSM-III, it was moved to the somatoform disorders, and diagnostic criteria were provided. In a revision of the classification (DSM-III-R), a duration of 6 months was added, and patients with delusional beliefs were excluded. The DSM-IV criteria are shown in Table 5.2.5.2. They exclude patients whose symptoms are better explained by other anxiety, depressive, or somatoform disorders.[1] Also, in DSM-IV, specific phobia of illness is separated from hypochondriasis. The illness phobic is said to fear contracting an illness whereas the hypochondriac fears disease already present.

The ICD-10 criteria for hypochondriacal disorder differ from those in DSM-IV. They require a persistent belief about having one or more specifically named serious physical diseases.[6] In addition, they include body dysmorphic disorder. With respect to illness behaviour, the ICD-10 criteria state that hypochondriacal concerns cause persons to seek medical investigation or treatment. They also state that patients may accept reassurance in the short-term, but that in the long run they are not likely to respond.

Table 5.2.5.2 Abbreviated DSM-IV diagnostic criteria for hypochondriasis

(a) Preoccupation with fears of having, or the idea that one has, a serious disease based on misinterpretation of bodily symptoms
(b) The preoccupation persists despite appropriate medical evaluation and reassurance
(c) Belief not of delusional intensity
(d) Preoccupation causes significant distress or impairment
(e) Duration of at least 6 months
(f) Not better accounted for by other anxiety, depressive, or somatoform disorders

The somatoform disorders category to which hypochondriasis belongs is controversial, and many question its inclusion in the classification.[7] They see these disorders as ill-defined, of questionable validity and based more on illness behaviour than on distinctive features. They also view them as creations of Western biomedicine that serve to devalue patients who challenge the theoretical model upon which it is based.[8] According to that model, illness is a response to disease, and the person who is ill without disease, e.g. hypochondriasis, is marginalized.

Were the somatoform disorders to be eliminated, some have proposed moving hypochondriasis to the anxiety disorders (health anxiety) or to a proposed grouping, the obsessive-compulsive spectrum disorders.

Validity

Evidence for the validity and utility of the diagnosis of hypochondriasis remains limited. In studies aimed at demonstrating validity, Barsky *et al.*[9] showed that distinguishing characteristics of the disorder aggregated in some medical outpatients but were less common in others. The same patients had other features of hypochondriasis indicating external validity. Using a structured interview for hypochondriasis, these investigators and others[10,11] observed a positive correlation between interview and physician ratings (concurrent validity). Hypochondriacal patients also had more ancillary features of hypochondriasis than did control patients (external validity). Also, other clinical characteristics distinguished interview positive from interview negative patients, indicating discriminate validity. Follow-up studies have shown a degree of diagnostic stability suggesting predictive validity.[12,13]

Measures

A variety of measures have been developed to screen for hypochondriasis and assess the severity of hypochondriacal concerns.[14] These are shown in Table 5.2.5.3. The Whiteley Index, a self-report instrument based on the observed characteristics of hypochondriacal psychiatric patients, is one of the most widely used.[15] It consists of 14 yes versus no items, but recent work suggests that a 7-item version is satisfactory for screening. The Illness Attitude Scales is a 27-item measure of psychopathology associated with hypochondriasis.[16] A principal components' analysis yielded two factors, one measuring health anxiety and the other illness behaviour. The health anxiety subscale has been used to distinguish hypochondriacal from non-hypochondriacal patients.

Recently, self-assessment measures have been developed to assess the various dimensions of health anxiety and hypochondriasis. The Health Anxiety Inventory contains 47 items covering a range of hypochondriacal features.[17] An advantage of this scale is that it distinguishes patients with high health anxiety from those with physical illness.

The Structured Clinical Interview for DSM-IV (SCID) and the Composite International Diagnostic Interview (CIDI) are comprehensive diagnostic interviews that contain somatoform disorder modules. The CIDI has been used in epidemiologic surveys. Its stem question for hypochondriasis is, 'In the past 12 months, have you had a period of 6 months or more when most of the time you worried about having a serious physical illness or deformity?'

Based on the SCID, Barsky *et al.*[10] developed a structured interview that focuses exclusively on hypochondriasis. It begins with a series of probe questions that, if answered affirmatively, trigger the remaining interview. It is suitable for confirming the diagnosis in a screened population.

Diagnostic assessment remains less than satisfactory because the threshold for caseness has not been established, medical and psychiatric comorbidity make diagnostic decision-making difficult, and independent medical evaluation is rarely part of the process.

Table 5.2.5.3 Measures for the assessment of hypochondriasis

Self-rated questionnaires
Whiteley index
Illness worry scale
Illness attitude scales
Health anxiety questionnaire
Health anxiety inventory
Multidimensional inventory of hypochondriacal traits
Psychiatric diagnostic screening questionnaire
Structured interviews
Structured diagnostic interview for hypochondriasis
Structured clinical interview for DSM-IV
Composite international diagnostic interview
Schedules for clinical assessment in neuropsychiatry

Differential diagnosis

Physical disorders

A few hypochondriacal patients suffer from undetected physical disease. Consequently, it is important to exclude medical conditions that, in their early stages, may cause vague symptoms with few signs or laboratory abnormalities. These include neurological conditions, such as multiple sclerosis or myasthenia gravis; endocrine conditions, such as thyroid or parathyroid disorders; multisystem disease such as systemic lupus erythematosus or occult malignancies. Because of such possibilities, a physical cause warrants continuing consideration even after the initial work-up has been completed.

Psychiatric disorders

Patients with **panic disorder** may be difficult to distinguish from those with hypochondriasis because they commonly have hypochondriacal features. A diagnosis of hypochondriasis should not be made if illness concerns are better accounted for by panic disorder. Patients with hypochondriasis tend to fear the long-term consequences of illness (such as cancer) whereas those with panic fear the immediate consequences of illness events (such as a heart attack); the former fear death, the latter dying. Also, those with hypochondriasis misinterpret a range of bodily sensations, whereas those with panic misinterpret the symptoms of autonomic arousal.

Hypochondriasis must be distinguished from **specific phobia, illness subtype**.[1] Patients with hypochondriasis are preoccupied with a disease they believe is already present, whereas illness phobics fear developing a disease they do not yet have. Illness phobic symptoms are triggered by external as well as internal cues. For instance, exposure to a person with the feared disease may elicit a fear response.

Hypochondriasis must be distinguished from **obsessive-compulsive disorder**. Patients with the latter often have intrusive thoughts about disease or contamination and rituals that involve

checking or reassurance-seeking. They differ from patients with hypochondriasis in having other obsessions and compulsions. Obsessive-compulsive patients tend to regard their ideas as senseless and resist them, whereas those with hypochondriasis regard them with conviction.

Hypochondriasis must also be distinguished from **generalized anxiety disorder** which is characterized by excessive worry about a number of areas. These may include health but other areas are generally involved as well. If worry is confined to illness, then a diagnosis of GAD should not be made. Patients with GAD tend to have health worries that are general, whereas those with hypochondriasis involve specific diseases such as cancer.

Hypochondriasis that develops during an episode of **major depression** and remits with treatment of the mood disturbance may be better accounted for by the depressive disorder. In that case, the patient is likely to focus concern upon the vegetative symptoms of depression and interpret these as irreversible loss of health. On the other hand, a diagnosis of hypochondriasis may be appropriate when hypochondriacal concerns are not confined to an episode of depression and are not focused on symptoms of the mood disorder.

Hypochondriasis and **somatization disorder** are both characterized by somatic symptoms. However, patients with hypochondriasis worry about the meaning of symptoms rather than the symptoms themselves. They are concerned about the consequences of serious illnesses rather than securing the gains of illness (e.g. sick role) as are patients with somatization disorder. Patients with hypochondriasis have an equal sex distribution whereas those with somatization disorder are predominantly women.

Hypochondriacal beliefs of a delusional nature may occur in patients with psychoses, but these patients usually have other psychotic features. However, delusions of disease may be the main or only manifestation of **delusional disorder, somatic type**. Such delusions may be bizarre or unrealistic, whereas the beliefs of patients with hypochondriasis are overvalued.

Epidemiology

Prevalence

The prevalence of hypochondriasis in the **general population** has not been established. Major surveys of psychiatric disorders have either excluded the somatoform disorders or identified few cases. For instance, Looper and Kirmayer[18] found that 6 per cent responded affirmatively to screening for illness worry, but only 0.2 per cent met full criteria for hypochondriasis according to a structured interview. Two studies that focused exclusively on somatoform disorders obtained higher estimates (4.5 and 7.7 per cent).[19, 20] Two other surveys focusing on illness worry found that half the respondents with such worry had the illness they worried about.[18, 21] Among such people it may be difficult to distinguish excessive from normal worry.

The prevalence of hypochondriasis among **primary care outpatients** had been examined in a number of studies. In a cross-national survey, Gureje *et al.*[22] noted that, if the criterion of failure to respond to reassurance were set aside, 2.2 per cent of patients qualified for this diagnosis and were as impaired as those meeting full criteria. In studies based on structured interviews, prevalence estimates have ranged from 2.2 per cent to 9.4 per cent.

Hypochondriasis may be prevalent in **medical specialty populations** where patients with functional disturbances are common. For instance, one survey found the disorder in 13 per cent of otolaryngology clinic patients. Also, hypochondriacal concerns are higher in patients with functional than with organic illnesses. For example, in one study higher hypochondriasis scores were obtained from patients with irritable bowel syndrome than from patients with organic gastrointestinal disease. Hypochondriacal concerns and health anxiety are especially high in patients with chronic pain.

High health anxiety is one of the factors shared by functional somatic syndromes in the general population.[23] However, it is not clear whether this represents a vulnerability factor or a consequence of unexplained symptoms.

Risk factors

Risk factors for unexplained somatic symptoms include female gender, older age, non-white race, less education, and lower income. With respect to hypochondriasis, few of these demographic factors appear important although findings have been inconsistent. The risk for men appears to be equal that for women. Some studies have shown persons with illness worry and hypochondriasis to be older and to have more physical illness. Two studies found them to have less education.

Comorbidity

Hypochondriacal patients in primary care have high levels of psychological as well as somatic symptoms. Strong positive correlations have been observed between hypochondriacal concerns and depressive ($r = 0.58$), anxiety ($r = 0.55$), and somatic symptoms ($r = 0.52$). In one study, the proportions of hypochondriacal and control patients, having one or more comorbid disorder, were 62 and 30 per cent respectively. Anxiety and depressive disorders accounted for most of the excess.

Family and twin studies

Taylor *et al.*[24] used a twin study to examine the genetic and environmental contribution to excessive health anxiety. After controlling for medical morbidity, which may be a source of health anxiety, they found that genetic factors accounted for 37 per cent of the variance in fear of disease and 10 per cent in disease conviction. For both dimensions the remainder of the variance (63 and 90 per cent respectively) was accounted for by non-shared environmental factors. These and other results suggest that some dimensions of health anxiety are moderately heritable. They also suggest that such anxiety is largely a learned phenomenon.

A family study compared the first-degree relatives of probands with and without hypochondriasis obtained from a general medicine clinic.[25] No difference in the frequency of hypochondriasis was found between these groups of family members. However, certain traits and attitudes, such as hostility, low agreeableness, and dissatisfaction with care, were significantly higher among the relatives of hypochondriasis probands. Such traits and attitudes may confer vulnerability to hypochondriasis and/or other somatoform disorders.

Morbidity and service utilization

Hypochondriasis is associated with impairment in physical functioning and work performance. Patients with this disorder view their health as worse, and experience more physical disability as well as impairment in occupational roles than patients without hypochondriasis.[10, 11] They use more medical services yet are less satisfied

with them than non-hypochondriacal patients. This increased utilization includes physician visits, laboratory tests, outpatient costs, and hospitalizations. Hypochondriacal patients tend to feel that their medical problems have not been thoroughly evaluated and as a consequence consult many physicians (i.e. doctor-shopping).

Hypochondriasis and health anxiety tend to be associated with increased symptom reporting and functional impairment, although the findings from various clinical populations have been inconsistent. For instance, hypochondriacal concerns are associated with higher disability and lower quality of life among patients with irritable bowel syndrome, chronic fatigue, and fibromyalgia.[26] One study found hypochondriasis the strongest predictor of pain due to osteoarthritis, and another showed high health anxiety predictive of abdominal pain 1 year later. Hypochondriasis was also a predictor of disability in patients with coronary artery disease. Consistent with these observations, hypochondriasis is associated with increased reporting of, and distress from, medication side effects.

Aetiology and pathogenesis

Personality

Hypochondriacal concerns are strongly related to the major personality dimension of neuroticism or negative emotionality.[27] Positive correlations between neuroticism and hypochondriacal concerns ranging from 0.4 to 0.5 have consistently been observed in non-clinical samples. Neuroticism refers to a tendency to experience and report negative emotions and overreact to stress. Persons high on this dimension are prone to find bodily sensations noxious and interpret them as signs of serious illness. Neuroticism may represent a vulnerability factor for hypochondriasis.

Certain personality traits may have more to do with difficult patient–doctor relationships than with hypochondriasis itself. Patients with hypochondriasis have been described as angry and mistrustful. Such characteristics might reflect negative emotions belonging to the domain of neuroticism or the negative pole of agreeableness, another of the major personality dimensions. They might also reflect obsessive-compulsive or masochistic personality traits observed in some patients.

Developmental factors

Childhood influences appear to be important in the development of hypochondriasis. Reports of **traumatic events during childhood**, including physical and sexual abuse, have been elicited more frequently from hypochondriacal than non-hypochondriacal patients. Although findings are preliminary, they are consistent with a literature linking childhood neglect and abuse to unexplained somatic symptoms in adults.

Childhood experience of illness may contribute to the development of hypochondriasis. For instance, Noyes *et al.*[28] obtained reports of serious illness or injury before age 17 from a third of adults with hypochondriasis. Similar findings from patients with hypochondriasis and somatization have been reported by others. Early illness may create a sense of physical vulnerability in susceptible individuals. Childhood exposure to serious illness or death of a family member or friend may do likewise.

Parental attitudes may also contribute to hypochondriasis. Excessive concern for a child's health or overprotection on the part of a parent may lead to anxiety about health as may special caretaking and rewards for illness. A child may also model exaggerated illness behaviour displayed by a parent. The importance of developmental factors—early adversity, experience of illness, over solicitous parents—suggest that hypochondriasis is in large measure learned behaviour.

Life events

Stressful life events appear to be related to increased reporting of physical symptoms and hypochondriasis, although there have been few studies. Events involving illness and death may have a specific role as the symptoms of hypochondriacal patients sometimes resemble those of family members who have been ill or died. In addition, illness events may give rise to hypochondriacal symptoms; 'cardiac neurosis' following myocardial infarction is an example. Transient hypochondriasis has been observed following medical illness in predisposed individuals.

Cognitive and perceptual factors

According to the cognitive-perceptual model, hypochondriasis is based on misinterpretation of bodily symptoms as signs of serious disease and on the experience of somatic sensations as intense, noxious, and disturbing.[29] In this model, the faulty attribution of innocuous sensations is the central defect. A number of studies have shown that, when symptoms are attributed to pathological processes, they become intensified. Such attribution may focus attention on symptoms thereby amplifying them. Misinterpretation of this kind may arise from cognitive schemata that were established through earlier experience with illness.

The tendency to experience bodily sensations as intense and disturbing has been termed somatosensory amplification. One study that used the Somatosensory Amplification Scale found a positive correlation of 0.56 between amplification and hypochondriasis. This finding suggests that individuals with hypochondriasis have a constitutionally lowered threshold for physical symptoms or that they have a heightened attentional focus and increased physiological arousal.

Evidence of physiological abnormalities was obtained by Gramling *et al.*[30] In a preliminary investigation, they observed physiological reactivity that distinguished women with hypochondriasis from those without. Hypochondriacal subjects had a higher mean heart rate and lower mean hand temperature during a cold pressor test compared to controls. These subjects terminated the test more frequently and rated it as more unpleasant than did controls.

Interpersonal factors

According to the interpersonal model, hypochondriasis is a form of care-eliciting behaviour that finds expression in physical complaints. Through unexplained somatic symptoms and expressions of illness worry, patients with this disorder seek emotional and interpersonal support from family members and physicians. Need for support of this kind arises from insecure attachment that originated in early relationships with caregivers. In a test of this model, Noyes *et al.*[31] found that hypochondriacal concerns among primary care patients were associated with various insecure attachment styles. These concerns were also associated with interpersonal problems and lack of reassurance from medical care.

Social and cultural factors

Social and cultural factors are important determinants of hypochondriasis. Throughout the world physical symptoms are common vehicles for the communication of distress. Somatic distress

gains the attention of **family and community** because it signals impairment in functioning that could alter social roles. Such distress not only calls forth caretaking but also obtains the sick role for those with acute illness. This social role with its privileges and responsibilities protects society from the disruptive effects of illness and promotes the return to health and social functioning of its members. Persons who are socially isolated or lacking in social support are more likely to manifest care-eliciting behaviour such as hypochondriasis.

Physicians play an important role in the development of hypochondriasis. They may make alarming statements or fail to provide reassurance that is based on thorough evaluation. In addition, they may order unnecessary tests, diagnose undetected disease, or treat injudiciously. They may add to concerns by failing to diagnose the psychiatric disturbance—hypochondriasis—telling patients instead that nothing is wrong. In doing this, they challenge and reject patients thereby contributing to suffering and alienation from the health care system.

Cultural attitudes may contribute to hypochondriasis. The American lifestyle, which emphasizes fitness and attractiveness, fosters preoccupation with health and encourages people to see their distress in terms of physical illness. There are, for example, cultural differences in the threshold for pain, pain tolerance, patterns of arousal, and physiological and behavioural responses to pain.

Course and outcome

Course

Hypochondriasis may begin at any age including childhood. The onset may be associated with stressful life events that in some instances involve illness. Some individuals develop hypochondriacal concerns transiently and others lastingly in reaction to physical illness. Among family medicine patients, those who became hypochondriacal a year after initial assessment were found to have had more illness worry and unexplained symptoms and to have rated their health as worse at baseline than non-hypochondriacal patients. Ambiguous symptoms or illness events may contribute to hypochondriacal concerns in patients so predisposed. Hypochondriasis appears to follow a chronic, fluctuating course.

Outcome

Follow-up studies show that, after their initial clinic visit, most patients with hypochondriasis improve. Still, a substantial proportion continue to meet criteria for the disorder and many more have persisting symptoms. For instance, among hypochondriacal general medicine patients, Noyes *et al.*[12] and Barsky *et al.*[13] found that, after 1 to 4 years, two-thirds continued to qualify for the diagnosis and the remaining one-third had persisting symptoms. Thus, despite improvement, the patients continued to be more hypochondriacal, more impaired, and more symptomatic than non-hypochondriacal patients.

Like patients in general, those with hypochondriacal concerns tend to seek care when they are most distressed. Their subsequent improvement may represent a natural fluctuation, a response to physician contact or to non-specific treatment. Some patients report having responded to reassurance. In a few instances, serious medical illness may relieve hypochondriacal concerns by legitimizing symptoms.

Studies indicate that greater severity and longer duration of symptoms are predictive of worse outcome. Failure to remit in one or more follow-up studies was predicted by more severe hypochondriacal concerns and somatic symptoms, longer duration of hypochondriasis, more psychiatric comorbidity, poorer perception of health, and greater neuroticism.

Complications

There is little information concerning complications of hypochondriasis. Because some patients utilize extensive medical care, one might expect complications resulting from repeated or unnecessary evaluations, tests, procedures, or treatments. Such iatrogenic complications have been reported for somatoform disorders but there is little documentation for hypochondriasis. On the other hand, physical illness may be overlooked in patients whose problems are considered psychiatric. There is almost no information on mortality. Suicide is said to be rare in hypochondriasis unless accompanied by severe depression in which case the risk may be increased.

Treatment

Until recently, the treatment of hypochondriasis was regarded with pessimism. It now appears that effective psychological, even pharmacological, interventions are being developed. A variety of approaches have been proposed but controlled trials of cognitive-behavioural therapy have established its efficacy, and preliminary trials of antidepressant medication have shown promise.

Psychological therapies

Most hypochondriacal patients, referred to mental health professionals, receive **psychotherapy** although such treatment has received little study. In one controlled trial a small number of patients with hypochondriasis were randomly assigned to explanatory therapy or a waiting list. The therapy yielded significant improvement in illness behaviour and health care utilization compared to no treatment, and gains were maintained for 6 months. This form of therapy involves repeated physical examinations, reassurance concerning symptoms, and information about psychophysiologic processes. Additional controlled trials of this and other forms of psychotherapy (e.g. psychodynamic, interpersonal) are clearly needed.

Four randomized, controlled trials for patients with hypochondriasis have shown that **cognitive behavioural therapies** are superior to no therapy with benefits sustained for up to 12 months.[32–35] These trials show that psychological treatment is efficacious for referred patients. However, one study showed that behavioural stress management, a non-specific intervention, was effective as well,[33] and another showed that cognitive and behavioural procedures, by themselves, were equally effective.[34]

Cognitive procedures include identifying and challenging dysfunctional thoughts and formulating more realistic beliefs. Behavioural procedures involve exposure *in vivo* with response prevention. These techniques include exposure to feared internal and external stimuli (e.g. physical exercise, visiting sick persons, reading about feared diseases, writing one's obituary) and prevention of checking and reassurance-seeking behaviours.

These trials showed that psychological treatment is effective but leave important questions unanswered. For instance, is psychological

therapy acceptable to most hypochondriacal patients in primary care? Are the techniques specific or do the benefits result from non-specific factors (e.g. therapeutic attention, therapist–patient relationship, credible procedures)? Also, are these treatments cost-effective? One trial involved up to 16 sessions over 4 months, which is expensive in terms of time and resources.

In consideration of these issues, several authors have advocated a **group approach**. For example, one study showed that group treatment is feasible. To improve acceptance, the authors referred to their intervention as a course in stress management and carried it out in a general practice setting.

Pharmacological therapies

There is evidence that patients with secondary hypochondriasis respond to drug therapy for the primary disorder. For example, Noyes *et al.*[36] assessed hypochondriacal concerns in patients receiving pharmacological treatment for panic disorder and agoraphobia. At the completion of treatment, a significant reduction in concerns was observed among those whose anxiety symptoms had improved. Observations of a similar kind have been made in patients with major depression.

No randomized controlled trials of pharmacotherapy for hypochondriasis have yet been completed, but a series of open label studies suggest that medication has promise. For example, Fallon *et al.*[37] reported that 10 or 16 patients with primary hypochondriasis given fluoxetine were very much improved after 12 weeks. And others have reported similar results with paroxetine, fluvoxamine, and nefazodone. Of the more than 50 patients enrolled in these trials, two-thirds responded to an SSRI. In these trials, drugs were relatively well tolerated and few patients dropped out because of side effects. This is noteworthy in view of the sensitivity to adverse effects observed in such patients. Controlled trials are needed to show proof of efficacy in primary hypochondriasis.

Management

Most hypochondriacal patients are best managed by their primary physicians. Few are successfully referred for specialty care because the focus of their concerns is, at least initially, on unexplained somatic symptoms. Although for some the ultimate goal is specific treatment, such treatment is not yet widely available.

Successful management depends upon a trusting relationship with a physician. To establish this, the physician should first **legitimize the patient's symptoms** by listening carefully and completing a thorough evaluation. Respectful treatment and statements to the effect that unexplained symptoms are nonetheless real are often helpful (see Table 5.2.5.4).

The scheduling of **regular visits** is an important strategy. Such visits serve several purposes. First, they reduce the reward for more severe or new symptoms that patients often present at unscheduled calls or visits. Next, they assure patients that the physician has an ongoing interest in their well-being. Finally, they provide reassurance through continued health monitoring.

Physicians should use **restraint in evaluating** hypochondriacal patients. New symptoms must be thoroughly evaluated, but overly aggressive diagnostic evaluation can be counterproductive. Extensive and dramatic tests can generate alarm, and when testing is repeated, it may convey physician uncertainty. Physicians should also avoid making diagnoses simply to have something to treat.

Physicians should also **approach treatment cautiously**. Medications, even when prescribed for benign indications, cause patients to worry about the conditions for which they are given. And too often they result in intolerable side effects and iatrogenic complications.

Hypochondriacal patients need an **explanation for their distress**, one that counters the notion of serious disease. Patients may be told their problem lies in the central nervous system processing of bodily sensations; this means they have a sensitive nervous system that amplifies discomforts and dysfunctions. Such an explanation gives legitimacy to the problem and avoids the stigmatizing label of hypochondriasis. Alternatively, patients may be told that they suffer from excessive health anxiety or worry. An explanation of the role of anxiety in altering attention, amplifying bodily sensations, and generating physiological symptoms may also be acceptable.

The goal of medical management is not to remove symptoms but help patients cope with them. The expectations of patients seeking elimination of symptoms may need to be modified. Reduced dependence on the technical aspects of care (namely, diagnostic testing and corrective intervention) is an important aspect of this overall objective. Patients need assistance in managing their lives so as to minimize continuing symptoms. The **aim is improved functioning**, a greater sense of control, and improved self-esteem. These objectives may accompany a gradual return to work and meaningful activity, and may be enhanced by improvements in exercise, diet, and daily routine.

Developing and **maintaining a therapeutic relationship** with the hypochondriacal patient is often challenging. The patient may be mistrustful and feel that his or her suffering is not understood. Masochistic and obsessional personality traits may contribute to a difficult doctor–patient relationship. A patient with such traits may seek mistreatment and thwart the physician's attempts to be helpful. Yet, a positive relationship is the key to successful management and can be achieved with acceptance, empathy, and understanding.[38]

Table 5.2.5.4 Management strategies for patients with hypochondriasis

Legitimize the patient's symptoms
Establish a regular schedule of visits
Base diagnostic evaluation on objective findings
Approach treatment of physical symptoms cautiously
Provide a plausible explanation for symptoms
Establish a goal of improved functioning

Specific treatment

Many hypochondriacal patients have psychiatric comorbidity, and treatment of comorbid anxiety and depressive disorders may yield significant improvement. If hypochondriasis has arisen during the course of an anxiety or depressive disorder, then successful treatment of the primary disorder may bring remission of hypochondriacal symptoms.

Specific treatments, to be acceptable, must be available in the primary care setting. Treating professionals must let patients know that their concerns are legitimate and their suffering understood. Beyond this, they must place a premium on engaging the patient, techniques for which have been described by a number of authors.

These patients are prone to drug side effects and often discontinue medication. For this reason, initial doses should be small, with gradual increases according to a modifiable schedule. Treating physicians should acknowledge the patient's sensitivity, and indicate that side effects are likely but may be dealt with.

Hypochondriasis is a significant medical condition for which treatment is now available.

Further information

Asmundson, G.J.G., Taylor, S. and Cox, B.J. (eds.) (2001). *Health anxiety: clinical and research perspectives on hypochondriasis and related disorders*. John Wiley, New York.

Lipsitt, D.R. and Starcevic, V. (eds.) (2001). *Hypochondriasis: modern perspectives on an ancient malady*. Oxford University Press, New York.

Taylor, S. and Asmundson, G.J.G. (2004). *Treating health anxiety: a cognitive-behavioral approach*. Guilford Press, New York.

References

1. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
2. Gillespie, R.D. (1928). Hypochondria: its definition, nosology, and psychopathology. *Guy's Hospital Report*, **8**, 408–60.
3. Creed, F. and Barsky, A. (2004). A systematic review of the epidemiology of somatization disorder and hypochondriasis. *Journal of Psychosomatic Research*, **56**, 391–408.
4. Starcevic, V. (2001). Clinical features and diagnosis of hypochondriasis. In *Hypochondriasis: new perspectives on an ancient malady* (eds. V. Starcevic and D.R. Lipsitt), pp. 21–60. Oxford University Press, New York.
5. Barsky, A.J., Ahern, D.K., Bailey, E.D., *et al.* (2001). Hypochondriacal patient's appraisal of health and physical risks. *American Journal of Psychiatry*, **158**, 783–7.
6. World Health Organization. (1993). The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. World Health Organization, Geneva.
7. Mayou, R., Kirmayer, L.J., Simon, G., *et al.* (2005). Somatoform disorders: time for a new approach in DSM-V. *American Journal of Psychiatry*, **162**, 847–55.
8. Rief, W. and Sharpe, M. (2004). Somatoform disorders—new approaches to classification, conceptualization and treatment. *Journal of Psychosomatic Research*, **56**, 387–90.
9. Barsky, A.J., Wyshak, G., and Klerman, G.L. (1986). Hypochondriasis: an evaluation of the DSM-III criteria in medical outpatients. *Archives of General Psychiatry*, **43**, 493–500.
10. Barsky, A.J., Cleary, P.D., Wyshak, G., *et al.* (1992). A structured diagnostic interview for hypochondriasis: a proposed criterion standard. *Journal of Nervous and Mental Disease*, **180**, 20–7.
11. Noyes, R., Kathol, R.G., Fisher, M.M., *et al.* (1993). The validity of DSM-III-R hypochondriasis. *Archives of General Psychiatry*, **50**, 961–70.
12. Noyes, R., Kathol, R.G., Fisher, M.M., *et al.* (1994). One-year follow-up of medical outpatients with hypochondriasis. *Psychosomatics*, **35**, 533–45.
13. Barsky, A.J., Fama, J.M., Bailey, D., *et al.* (1998). A prospective 4- to 5-year study of DSM-III-R hypochondriasis. *Archives of General Psychiatry*, **55**, 737–44.
14. Speckens, A.E.M. (2001). Assessment of hypochondriasis. In *Hypochondriasis: new perspectives on an ancient malady* (eds. V. Starcevic and D.R. Lipsitt), pp. 61–88. Oxford University Press, New York.
15. Pilowsky, I. and Spence, N.D. (1983). *Manual for the illness behavioural questionnaire* (*IBQ*) (2nd edn.). Department of Psychiatry, University of Adelaide, Adelaide, South Australia.
16. Kellner, R. (1986). *Somatization and hypochondriasis*. Praeger, New York.
17. Salkovskis, P.M., Rimes, K.A., Warwick, H.M.C., *et al.* (2002). The health anxiety inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. *Psychological Medicine*, **32**, 843–53.
18. Looper, K. and Kirmayer, L.J. (2001). Hypochondriacal concerns in a community population. *Psychological Medicine*, **31**, 577–84.
19. Faravelli, C., Salvatori, S., Galassi, F., *et al.* (1999). Epidemiology of somatoform disorders: a community survey in Florence. *Social Psychiatry and Psychiatric Epidemiology*, **32**, 24–9.
20. Noyes, R., Happel, R.L., and Yagla, S.J. (1999). Correlates of hypochondriasis in a nonclinical population. *Psychosomatics*, **40**, 461–78.
21. Noyes, R., Carney, C.P., Hillis, S.L., *et al.* (2005). Prevalence and correlates of illness worry in the general population. *Psychosomatics*, **46**, 529–39.
22. Gureje, O., Üstün, T.B., and Simon, G.E. (1997). The syndrome of hypochondriasis: a cross-national study in primary care. *Psychological Medicine*, **27**, 1001–10.
23. Aggarwal, V.R., McBeth, J., Zakrzewska, J.M., *et al.* (2006). The epidemiology of chronic syndromes that are frequently unexplained: do they have common associated factors? *International Journal of Epidemiology*, **35**, 468–76.
24. Taylor, S., Thordarson, D.S., Jang, K.L., *et al.* (2006). Genetic and environmental origins of health anxiety: a twin study. *World Psychiatry*, **5**, 47–50.
25. Noyes, R., Holt, C.S., Happel, R.L., *et al.* (1997). A family study of hypochondriasis. *Journal of Nervous and Mental Disease*, **185**, 223–32.
26. Robbins, J.M., Kirmayer, L.J., and Kapusta, M.A. (1990). Illness worry and disability in fibromyalgia syndrome. *International Journal of Psychiatry in Medicine*, **20**, 49–63.
27. McClure, E.B. and Lilienfeld, S.O. (2001). Personality traits and health anxiety. In *Health anxiety: clinical and research perspectives on hypochondriasis and related conditions* (eds. G.J.G. Asmundson, S. Taylor, and B.J. Cox), pp. 65–91. John Wiley, New York.
28. Noyes, R., Stuart, S., Langbehn, D.R., *et al.* (2002). Childhood antecedents of hypochondriasis. *Psychosomatics*, **43**, 282–9.
29. Salkovskis, P.J. and Warwick, H.M.C. (2001). Meaning, misinterpretations, and medicine: a cognitive-behavioural approach to understanding health anxiety and hypochondriasis. In: *Hypochondriasis: modern perspectives on an ancient malady* (eds. V. Starcevic and D.R. Lipsitt), pp. 202–22. Oxford University Press, New York.
30. Gramling, S.E., Clawson, E.P., and McDonald, M.K. (1996). Perceptual and cognitive abnormality model of hypochondriasis: amplification and physiological reactivity in women. *Psychosomatic Medicine*, **58**, 423–31.
31. Noyes, R., Stuart, S. Langbehn, *et al.* (2003). Test of an interpersonal model of hypochondriasis. *Psychosomatic Medicine*, **65**, 292–300.
32. Warwick, H.M.C., Clark, D.M., Cobb, A.M., *et al.* (1996). A controlled trial of cognitive behavioral treatment of hypochondriasis. *British Journal of Psychiatry*, **169**, 189–95.
33. Clark, D.M., Salkovskis, P.M., Hackman, A., *et al.* (1998). Two psychological treatments for hypochondriasis: a randomized controlled trial. *British Journal of Psychiatry*, **173**, 218–25.
34. Visser, S. and Bouman, T.K. (2001). The treatment of hypochondriasis: a randomized controlled trial. *Behavior Research and Therapy*, **39**, 423–42.
35. Barsky, A.J. and Ahern, D.K. (2004). Cognitive behavioral therapy for hypochondriasis: a randomized controlled trial. *Journal of the American Medical Association*, **291**, 1464–70.

36. Noyes, R., Reich, J., Clancy, J., *et al.* (1986). Reduction in hypochondriasis with treatment of panic disorder. *British Journal of Psychiatry*, **149**, 631–5.
37. Fallon, B.A., Liebowitz, M.R., Salman, E., *et al.* (1993). Fluoxetine for hypochondriacal patients without major depression. *Journal of Clinical Psychopharmacology*, **13**, 438–41.
38. Starcevic, V. (2002). Overcoming therapeutic pessimism in hypochondriasis, *American Journal of Psychotherapy*, **56**, 167–77.

5.2.6. Pain disorder

Sidney Benjamin and Stella Morris

Introduction

Persistent somatoform pain disorder is an ICD-10 diagnosis, which is included in the group of somatoform disorders. The term **pain disorder** is used in DSM-IV, and for convenience that is the term used here to refer to both classifications, unless a distinction needs to be made. This chapter aims to clarify the relationship of pain to mental disorders, the diagnosis of pain disorder and its differential diagnosis, and then considers how psychosocial factors contribute to pain, the treatments that stem from them, and the psychiatrist's potential contribution.

Pain has been defined by the International Association for the Study of Pain (**IASP**) as 'an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage'. 'Pain' is used here in this sense; it is not used primarily to indicate mental distress or anguish. As a perception, pain is essentially a subjective experience, and is directly accessible only to the patient. By contrast, tissue damage can be assessed by others, and its relationship with the subjective characteristics of pain have been shown to be variable, modulated by social and cultural experience, as well as within the central and peripheral nervous system.

Pain and the psychiatrist

Psychiatrists are likely to see patients with pain in psychiatric, general hospital, and community settings. Pain is associated with a wide range of mental disorders, and there are different ways in which this relationship may arise.

Pain may contribute to the cause of a mental disorder; for example, when a patient with cancer has pain, which is unrelieved by analgesics, and becomes depressed. This can result in additional distress and disability, and subsequently an exacerbation of pain. Treatment of depression may contribute to the relief from pain and improve the quality of life.

In a general hospital psychiatrists may see patients with **acute pain**, like the patient described above, but more often will see patients with **chronic pain**. Whatever the initial cause, the longer pain persists the more likely is it to result in the development of inappropriate patterns of illness behaviour and to have a profound effect on relationships with the family and other carers, presenting more complex challenges for management and poorer prognosis.

Pain disorder

Diagnostic and clinical features

Persistent somatoform pain disorder in **ICD-10** is the only somatoform disorder that is essentially characterized by pain. The **diagnostic requirements** are as follows:

1 'persistent, severe, and distressing pain';

2 pain 'cannot be explained fully by a physiological process or a physical disorder';

3 'pain occurs in association with emotional conflict or psychosocial problems that are sufficient to allow the conclusion that they are the main causative influences'.

There are also likely to be many of the features that occur in the other somatoform disorders, which have been described in previous chapters. The pain can be localized, as in low back pain, or generalized, as in fibromyalgia.

In **ICD-10**, the diagnosis is excluded if pain, presumed to be mainly psychological in origin, occurs in the course of schizophrenia or depressive disorder, or is believed to be due to psychophysiological mechanisms such as muscle tension. The main **differential diagnosis**, according to ICD-10, is the histrionic elaboration of pain primarily due to organic causes, particularly if this has not yet been diagnosed. In practice, it is uncommon for pain that has been properly investigated, and has persisted for more than 6 months, to be found subsequently to have a specific organic cause.

The **DSM-IV** diagnosis of '**pain disorder**' also needs to be considered because the requirements for diagnosis and the underlying rationale are rather different. This diagnosis is divided into three **subtypes**:

1 '**Pain disorder associated with psychological factors**', in which psychological factors are judged to play the major role, and physical disorders play either no part or only a minor part in its onset or maintenance.

2 '**Pain disorder associated with both psychological factors and a general medical condition**', in which both psychological processes and an organic disorder are judged to make important contributions to causation.

3 '**Pain disorder associated with a general medical condition**', due to an organic disorder and in which psychological factors are judged to make no contribution or to play only a minor role. This subtype is not regarded as a mental disorder but is coded on Axis III.

For the first two subtypes the diagnostic criteria, all of which must be satisfied, are summarized as follows:

(a) Pain, localized or more general, is the predominant symptom and its severity warrants clinical attention

(b) Pain results in distress, and impairment in social, occupational, or other areas of functioning.

(c) Psychological factors are judged to have an important role in the onset, severity, exacerbation, or maintenance of pain.

(d) It is not intentionally produced or feigned (factitious disorder and malingering are specifically excluded).

(e) Pain is not better accounted for by a mood, anxiety, or psychotic disorder and does not meet criteria for dyspareunia.

Pain disorder can also be coded according to whether it is acute or chronic (less or more than 6 months duration).

Comparison of ICD-10 and DSM-IV

The diagnoses of pain disorder in ICD-10 and DSM-IV share a number of characteristics. Pain disorder should be diagnosed as a mental disorder if psychological factors are thought to make a significant contribution to predisposition, precipitation, or maintenance, or to the severity of pain. In ICD-10, there should be evidence that emotional conflict or psychosocial problems are the main 'causative influences', whereas in DSM-IV psychological factors are judged to play either the 'major role' or 'an important role'. In both, the diagnosis can be made even though there may be possible or definite evidence of an organic disorder that contributes to pain (for instance, a prolapsed intervertebral disc), provided that this is judged to be insufficient to account fully for the features of pain. Both classifications stress the severity of pain and the distress caused by it, but only DSM-IV specifically requires a degree of disability as a diagnostic feature. The implication is that diagnosis requires detailed physical and psychiatric evaluation, including an assessment of the family and social context, as well as of disability.

Differential diagnosis of pain disorder

Pain can occur in the setting of virtually any mental disorder. Table 5.2.6.1 lists the ICD-10 diagnoses and their DSM-IV equivalents in which pain may be a predominant feature. The general description of most of these disorders is provided in other chapters of this book and the following account focuses only on aspects relevant to pain.

(a) Organic disorders

Many painful disorders have a well-recognized organic pathology that accounts for the occurrence of pain (for example, angina, sickle cell arthropathy), but psychosocial processes tend to modify the severity of pain and associated disability. Thus, psychological and social interventions may make an important contribution to management, and as pain becomes more chronic, or fails to respond to usually effective physical treatments, psychosocial interventions assume greater significance. These disorders can be diagnosed in ICD-10 within the diagnoses headed '**Psychological interactions with physical disorders**' in Table 5.2.6.1.

(b) Pain syndromes of uncertain origin

There are many disorders characterized by pain, which are essentially syndromes with no known consistent organic pathology (Table 5.2.6.2). Psychological and social factors are thought to contribute to the development and maintenance in many cases,[1] but psychological causes specific to these different syndromes have not been identified. Patients with these pain syndromes tend to have a greater prevalence of non-psychotic mental disorders than is found in the general population. The pain itself can usually be accommodated in ICD-10 within the categories of **somatoform autonomic dysfunction** or **somatoform pain disorder** (see below). The 'diagnoses' listed in Table 5.2.6.2 tend to be used by non-psychiatrists to describe clusters of medically unexplained symptoms and are terms which are likely to be acceptable to patients. Treatments for these disorders generally include physical approaches, often of limited efficacy, as well as a range of psychosocial interventions, which are described below.

(c) Pain and mental disorders

(i) Psychoses

At the beginning of the twentieth century, French psychiatrists described **coenestopathic states** as disorders characterized by unpleasant sensations, particularly pains, thought to be of central origin, but unrelated to organic brain disease.[2] Such disorders were a daily occurrence in psychiatric clinics, commonly associated with the psychoses, and in this setting were related to **somatic hallucinations** and **systematized delusional states**. Such presentations are now described infrequently in Europe and North America.

Patients with any psychosis may complain of pain, sometimes with bizarre descriptions of quality and **delusional attribution**. In practice, it is difficult to differentiate between a **somatic hallucination** and an illusion (arising from physiological or pathological processes). Complaints of pain in psychotic disorders have no psychiatric diagnostic specificity. Pain has been described particularly in association with schizophrenia and depressive psychoses, but may occur in any psychotic disorder. In the course of a psychotic disorder, illusions and delusional interpretations of pain may arise from unrelated organic disorders and therefore require careful **physical assessment**.

(ii) Mood- and anxiety-related disorders

These are by far the most common mental disorders associated with pain in most settings. In the general population, 12 per cent of adults have experienced **chronic widespread pain** (defined according to the criteria of the American College of Rheumatologists) in the previous 3 months and their prevalence of mental disorders is three times that of the pain-free population.[3] Most of these diagnoses are mood and anxiety disorders, with the former being more common in those with chronic pain. In **pain clinic settings**, the prevalence of mental disorders varies according to referral patterns, but about 30 to 40 per cent of patients have depressive disorders, and this is similar in those with and those without a relevant physical disorder.[4] Those without organic disorders tend to have lower ratings for both mood disorders and pain severity. Those with mood disorders report more severe pain.

Diagnosis of mood and anxiety disorders is based on the usual standardized criteria, but may be missed due to the process of **somatization**, particularly where patients attribute their depressed mood to pain and an underlying physical condition (whether present or not) and invite their doctors to share this belief.[5] In the past, pain has been thought of as a proxy for depression, giving rise to the concept of a '**depressive equivalent**' or '**masked depression**'. This has been based mainly on evidence for the psychogenicity of chronic pain rather than a specific relationship to depressive disorders, has received widespread criticism, and has not advanced theoretical knowledge or clinical practice.

(iii) Post-traumatic stress disorder

Many patients with post-traumatic stress disorder (PTSD) have been subjected to actual or threatened physical injury, so it is not surprising that pain is one of the commonest symptoms that they report, the prevalence ranging from 20 to 80 per cent. Further, 10 to 50 per cent of patients with chronic pain satisfy criteria for PTSD, and patients with musculoskeletal pain are four times more likely to develop PTSD than those without it.[6] **Pain disorder and PTSD** can be diagnosed jointly, if criteria for both are satisfied. Mechanisms including shared vulnerability, fear-avoidance, and

Table 5.2.6.1 Mental disorders included in the differential diagnosis of pain disorder

ICD-10		DSM-IV	
Psychotic disorders			
F00–09	Organic mental disorders	290	Dementia 293 Delirium
F20–29	Schizophrenia, schizotypal, and delusional disorders	273	Schizophrenia and other psychotic disorders
Mood- and anxiety-related disorders			
F32/33	Depressive episode	296.2/3	Major depressive disorder
F34.1	Dysthymia	300.4	Dysthymic disorder
F41	Anxiety disorders	300.02	Generalized anxiety disorder
F43.1	Post-traumatic stress disorder	309.81	Post-traumatic stress disorder
F43.2	Adjustment disorders	309	Adjustment disorders
Somatoform disorders			
F44.4	Dissociative (conversion) disorders	300.11	Conversion disorder
F45.0	Somatization disorder	300.81	Somatization disorder
F45.1	Undifferentiated somatoform disorder	300.81	Undifferentiated somatoform disorder
F45.2	Hypochondriacal disorder	300.7	Hypochondriasis
F45.3	Somatoform autonomic dysfunction	300.8	Pain disorder
F45.4	Somatoform pain disorder	300.81	Somatoform disorder NOS
F45.8	Other somatoform disorders		
F45.9	Somatoform disorder, unspecified		
Other neurotic disorders			
F48.0	Neurasthenia		
F48.8	Other specified neurotic disorders (occupational neurosis, e.g. writer's cramp)		
Sexual disorders			
F52.5	Non-organic vaginismus	306.51	Vaginismus
F52.6	Non-organic dyspareunia	302.76	Dyspareunia
Psychological interactions with physical disorders			
F54	Pyschological or behavioural factors associated with disorders or diseases classified elsewhere	316	Psychological factors affecting medical condition
F68.0	Elaboration of physical symptoms for psychological reasons		
Disorders of behaviour			
F68.1	Intentional production or feigning of symptoms	300.19	Factitious disorder
		V65.2	Malingering
Comorbidity of pain disorder			
Any of the above except psychoses and other somatoform disorders			
Substance abuse			
F10	Disorders due to alcohol	291 & 303.9	Alcohol-induced disorders and dependence
F11–13	Disorders due to psychoactive substance abuse	292 & 304	Other substance-induced disorders and dependence
F55	Abuse of non-dependence-producing substances		
Personality disorders			
F60–62	Personality disorders and changes	301	Personality disorders

mutual maintenance have been postulated to account for this comorbidity.[7] This has implications for assessment (described below), and treatment programmes may need to be modified accordingly.

(iv) Somatoform disorders

Somatoform disorders are uncommon in people with chronic pain in the general population.[3] Prevalence varies considerably in clinical samples, but somatoform disorders have been reported in 12 to 52 per cent of patients,[4] so they include highly selected samples.

Complaints of pain occur commonly in each of the somatoform disorders and may be the predominant symptom. Multiple physical complaints, often including pains at different sites, fluctuate from time to time usually for many years, providing a characteristic feature of **somatization disorder**. In **hypochondriacal disorder** pain is a common complaint, and forms the focus for concern and overvalued beliefs about unidentified disease.

The diagnosis of **somatoform autonomic disorder** is based on autonomic arousal (palpitation, sweating, tremor), which must be a prominent feature of the clinical picture, together with physical complaints, often pain, referred to specific organs, systems, or parts of the body. As with other somatoform disorders, the patient will be distressed about the possibility of underlying physical disease and is not reassured by negative findings on appropriate assessment and explanation. This diagnosis is sometimes appropriate for syndromes listed in Table 5.2.6.2.

Pain, as a form of **conversion**, has a traditional place in the literature on **hysteria**, based on the concepts of psychogenicity, the

Table 5.2.6.2 Disorders of uncertain origin, presenting primarily with pain, in which psychosocial factors are thought to contribute to predisposition, precipitation, or course

Generalized
Fibromyalgia
Relatively localized
Tension headache—acute or chronic
Temporomandibular pain and dysfunction syndrome
Atypical facial pain
Atypical (non-cardiac) chest pain
Abdominal pain of psychological origin
Non-ulcer dyspepsia
Irritable bowel syndrome
Chronic pelvic pain
Irritable bladder syndrome
Procatalgia fugax

contribution of stressful experiences with dissociation, and primary gain. In recent years, however, research has focused on other psychological processes, and the concept of conversion as a primary mechanism now seems to be of limited interest. The category of **dissociative (conversion) disorder** in ICD-10 specifically includes sensory loss but excludes pain (sensory amplification), which therefore should not be diagnosed as a dissociative disorder. DSM-IV also excludes pain from the diagnosis of conversion disorder, unless other diagnostic criteria are satisfied.

The uncertain relationship and limited value of the different diagnoses included within the group of somatoform disorders in ICD-10 have been discussed in Chapter 5.2.1, and are well illustrated by the fact that pain may be a prominent feature of each category. Somatoform disorders presenting with pain are usually diagnosed as **somatization disorder or pain disorder**, with the former taking precedence if the diagnostic criteria are satisfied.

Comorbidity

Any physical or mental disorder may be diagnosed in addition to pain disorder. Anxiety and depression are common, and an additional diagnosis of **anxiety disorder** or **mood disorder** can be made if the criteria are satisfied. This dual diagnosis can be useful if, for example, a depressive disorder develops in the presence of a long-standing pain disorder. Any temporal relationship can occur, however, with pain onset preceding, developing simultaneously with, or following the onset of a mood disorder.

Other common comorbid diagnoses include **substance abuse** and dependence, sometimes of iatrogenic origin, and their management is an important component of pain-treatment programmes.[8] **Personality disorders** are an additional category of comorbidity. No single disorder predominates but histrionic, narcissistic, anxious (avoidant), and dependent features are all common in clinical practice, and anankastic traits may feed an inflexible focus on physical illness.

Epidemiology

Although the association of psychiatric symptoms with chronic pain has been studied in the general population, the prevalence of pain disorder, and other mental disorders presenting with pain, is uncertain because large-scale surveys of mental disorders do not include an assessment of pain and of related physical conditions.

Assessment of pain

Clinical assessment

The psychiatric assessment requires a full **psychiatric history** and **mental state examination**, with particular attention to those additional features relevant to pain. **The pain history** should include total duration (often underestimated by the patient), a detailed inquiry about the location and distribution of pain, including direct questions aimed at a total body survey, and the timing of first onset, subsequent periods of relapses and remissions, and their relationship to life events and difficulties. The **family history** should include assessment of severe, chronic or disabling physical disorders, and the patient's involvement with them. The **personal history** should include adverse childhood experiences (discussed below) and the **past history** of physical disorders and disability is particularly important.

Patients who **somatize** will tend to deny concurrent psychosocial events and their significance. For example, one of our patients was consistently unable to recall any distressing events in the year prior to the onset of severe, persistent, and disabling headache. His wife gave an account of the deaths of his father, brother, and closest friend during that year, and moreover described him as so distressed by these bereavements that he felt unable to attend any of the funerals. It is essential to take a **history from other informants**, and this can also provide an opportunity to assess the attitudes, knowledge, and beliefs of carers, and their interaction with the patient.

The patient's **pain beliefs and behaviours** (described below) are key aspects of the **mental state examination**. Patients often attribute chronic pain to an organic disorder and offer diagnoses; it is essential to review their **medical records** to assess the clinical findings and investigations, and the extent to which they support any diagnosis which is offered. Chronic pain associated with an underlying organic disorder may be exacerbated when the patient suffers a stressful life event, so it is important to **avoid assumptions of a dichotomy** of either **'organic' or 'psychogenic' pain**.

Standardized psychometric assessments

Many standardized questionnaires have been developed for the assessment of patients with chronic pain. They can be valuable for identifying mechanisms that contribute to pain, planning treatment, and monitoring changes during and after treatment. The evaluation of pain and associated beliefs and behaviours requires measures developed specifically for this purpose, and these are described below.

Other assessments, for example of **mood**, **illness behaviour**, and **social dysfunction**, have been developed within the field of pain research. Some measures are rather idiosyncratic, with uncertain psychometric properties, aimed at restricted diagnostic groups and clinical settings. This undermines the need to use consistent methods that allow comparison of different groups of patients, with physical, mental, and mixed disorders, at different places and times.

(a) Pain

The **severity** of pain can be assessed[9] using standardized **visual analogue scales** and **numeric analogue scales**. Such scales may have anchor points ranging from 'no pain' to 'the worst possible pain'.

The **quality** of pain can be assessed with **verbal descriptor scales**.[9] Factor analysis has resulted in the emergence of two that have best survived the test of time: an '**affective**' **dimension** (represented by words such as exhausting, terrifying, vicious), and a '**sensory**' **dimension** (e.g. stabbing, crushing, burning). They have been found consistently when administered in different languages and to different cultural groups. Ratings on both these scales are positively correlated with pain severity and mood ratings and, in the presence of mental disorders, contribute little to diagnosis.

The **topographical distribution of pain** can be assessed by using outline drawings of the body (front, back, and sometimes sides), which the patient is asked to shade to indicate the distribution of pain. These can help to identify pain that does not conform to physiological distributions and also widespread pain. Measures of pain intensity, quality, and distribution can be used toget her to capture the rather elusive and entirely subjective experience of pain.

(b) Pain behaviours

Although the experience of pain is entirely personal, it may be communicated to others by a range of verbal and non-verbal behaviours, which in some cases may be maladaptive, and which in turn influence the responses of others. Using a **learning theory model**, Fordyce[10] classified all pain into **'operant' and 'non-operant' pain**. The former includes all pain that is modified by positive or negative reinforcement, whether or not organic pathology is present. Standardized **structured assessments** are available to measure a range of well-defined **behaviours**.[11] These may include complaints of pain, requests for medication, groaning, facial expression, restricted mobility and the use of aids, time spent resting, and postures such as guarding and bracing. Such behaviours have been shown to fluctuate in response to changes in the environment, including different **attitudes and responses of carers**. This has led to the assessment of pain behaviours and their **environmental reinforcers**, and the development of pain-treatment programmes that originally focused on behavioural change by modifying reinforcement. Recent interest has focused on **pain-related fears** (e.g. of exacerbating pain by injury) and the management of consequent **avoidance**.[12]

(c) Pain beliefs

The belief that chronic as well as acute pain signals an underlying physical disease, which requires and should respond to physical intervention, whilst avoiding usual activities and functions, contributes to the development and maintenance of chronic pain and non-adherence to treatment, and the widespread dissatisfaction often expressed by patients and their doctors. Inappropriate beliefs that are relevant to pain assessment fall into three groups[13]:

1 beliefs about the nature of reality—for example, 'life should be pain-free';

2 beliefs in response to challenging circumstances, such as pain—including **locus of control**, **attributional style**, **cognitive errors**, and **coping strategies**;

3 specific ideas about the cause of a pain, appropriate management, and outcome.

The questionnaire assessment of pain-related beliefs has assumed increasing importance in the field of pain research,[13] with the recognition that pain beliefs interact with pain, cognitions, behaviours, affects and disability, and contribute to the prediction of outcome. Thus cognitive approaches to treatment are often integrated with behavioural management.

Psychosocial contributions to the development of pain

The origins of chronic pain are, in several respects, similar to those of somatization and other somatoform disorders. Current models of causation involve the interaction of biological, psychological, and social factors, each contributing to predisposition, precipitation, and maintenance.

The **family and personal histories** of patients with chronic pain include an excess of mood disorders, pain and disability, substance abuse, and personality disorders. Engel[14] described the dynamics of 'the **pain prone patient**' involving abusive childhood experiences, and noted how pain can become a pathway for the expression of guilt and expiation. Recent research[15] has reconsidered the significance of reports of **physical and sexual abuse** and other **adverse childhood experiences**. The relationship between chronic pain in adults and these childhood experiences appears, at least to some extent, to be determined by **selective reporting**, particularly in those with associated mental disorders,[16] but these experiences may make a significant contribution to pain in some individuals.

Precipitation of chronic pain is, in effect, **transition from acute to chronic pain**, and factors associated with this transition[17,18] include current mood and anxiety disorders, negative life events including physical illnesses and trauma, the social support network and dissatisfaction with work. A population based prospective study[19] found that new episodes of chronic widespread pain were predicted by the number of previous non-pain somatic symptoms and by a measure of illness behaviour which assessed numbers of consultations, treatments and perceived disability, and these two measures had an additive effect. Recent research has indicated the potential value of interventions designed to prevent the progression from acute to chronic pain.[17,18]

The psychiatric and psychological management of pain

Treatment of mental disorders

The treatment of chronic pain has presented a challenge to the ingenuity of health professionals, particularly because no single specialty or profession has the range of skills that is required. The treatment of mental disorders,[5] such as depressive or anxiety disorders,[20] is similar in most respects, whether or not pain is a prominent feature. In the presence of pain, however, mental disorders tend to be missed, and when recognized are treated inadequately. Depressive disorders with features indicating a good response to **antidepressants** should be treated with full therapeutic doses, but not with **narcotics**. **Anxiolytic drugs** including **benzodiazepines**, which result in dependence, should not be used in the treatment of these chronic disorders.

The use of antidepressants for pain relief

Antidepressant drugs are often used for the treatment of pain in patients who are not depressed. Randomized controlled trials[21] indicate that antidepressants, in doses within the usual therapeutic

range, provide more effective **analgesia** than placebo preparations in the treatment of diabetic neuropathy, postherpetic neuralgia, and atypical facial pain, as well as chronic non-malignant pain. Different **tricyclic antidepressants** (TCAs) appear to be equally effective and are more effective than **selective serotonin-reuptake inhibitors**. Data on **seretonin noradrenaline-reuptake inhibitors** are increasing and suggest that they may be effective, and preferable to TCAs because of a superior side effect profile.[22] The analgesic effect of antidepressants occurs in patients who are not depressed and is independent of any antidepressant effect.

Psychological treatments

Psychological treatments[5,23] are derived from different theoretical formulations of the aetiology of chronic pain. These include behavioural, cognitive, and psychodynamic approaches. Reviews of randomized controlled trials of **behavioural and cognitive approaches**[24] that have been developed specifically for the treatment of chronic pain illustrate the problems in assessing outcome due to different sampling methods, different types of control groups, non-standardized treatment components, and the different assessments that are included. Despite these limitations, the best studies demonstrate that these treatments are more effective than 'usual' medical treatment, remaining on a waiting list, or exercise programmes, and improvements can be sustained during lengthy follow-up periods.

Other approaches include various forms of 'stress management' including **relaxation techniques**, **biofeedback**, and **hypnosis**. Their value is uncertain; although pain ratings tend to be reduced, this is not a consistent finding on all measures.

Psychological treatments are rarely used in isolation, either from each other or from additional interventions, and integrating different approaches may enhance their effects.[23]

Multi-disciplinary pain management clinics

Pain-treatment centres[25] have been established in many countries and provide a diverse range of professional skills, treatments, and models of service delivery. In some, management is based mainly on anaesthetic techniques and medication, but psychological approaches are provided in others by **clinical psychologists** and **nurse therapists**. The management of problems due to inappropriate medication and **substance abuse** is an essential component of treatment.[8] Many clinics offer **structured programmes** of education and rehabilitation, with increments of **exercise**, to overcome **disability**, rather than aiming primarily at pain relief, and to which **physiotherapists or occupational therapists** may contribute. There is often an emphasis on the patient assuming increasing responsibility, rather than maintaining **dependence on medical services**.

Direct input from **psychiatrists** is variable and some centres specifically exclude the treatment of patients with serious mental disorders because their response to treatment is less certain. Although it is well recognized that social and environmental factors contribute to chronic pain problems, and can undermine progress following treatment, few specialized centres involve **carers** routinely in treatment or offer **family therapy**.

A range of physical, psychological, and social approaches should be offered, based on an **individual structured assessment** of needs. Members of the multi-disciplinary team require specific **training** in the management of pain. The work of the team has to be carefully coordinated, both within the team and with other health professionals, to avoid any ambiguity concerning the methods and goals of treatment.

Many pain clinics provide a treatment package in which cognitive therapy and graded exercise are predominant features. A similar approach is used for a number of other conditions, including **somatization disorder**, **hypochondriacal disorder**, **fibromyalgia**, and **chronic fatigue**, but the extent to which they may have similar origins and outcomes is uncertain.

Effects of treatment

The **outcome of psychological and psychiatric treatment** has been studied extensively,[5,24] but is difficult to evaluate because reports differ with regard to the characteristics of patients and disorders, inclusion criteria, assessments, and treatments as well as details of treatment delivery, attrition rates, choice of control groups, and the duration of follow-up. Many patients with chronic pain are unwilling to accept treatment and others are considered unsuitable. Nevertheless, psychological and rehabilitation treatments can have a sustained effect, based on the range of assessments that have been described. In addition, they can result in reductions in **sickness and benefit payments**,[26] **return to work**[27] and reduced **use and costs of medical services**.[28]

The outcome for patients with different mental disorders has not been assessed systematically. Patients involved in seeking **compensation** tend to have a poorer outcome, even after **litigation** has been concluded, but they can also benefit from treatment. There is some evidence that **secondary prevention** programmes may help to avoid the transition from acute to chronic pain in those who are particularly vulnerable.

Further information

For more information on the topic of this chapter, we have marked with an asterisk (*) those references, which will be of particular interest to the reader.

References

1. *Kellner, R. (1991). *Psychosomatic syndromes and somatic symptoms*. American Psychiatric Press, Washington, DC.
2. Dupré, E. (1913). *Les cénestopathies*. Reprinted as: Coenestopathic states. In *Themes and variations in European psychiatry* (eds. S.R. Hirsch and M. Shepherd), pp. 385–94. John Wright, Bristol, 1974.
3. Benjamin, S., Morris, S., McBeth, J., *et al.* (2000). The association between chronic widespread pain and mental disorder. *Arthritis and Rheumatism*, **43**, 561–7.
4. Benjamin, S., Barnes, D., Berger, S., *et al.* (1988). The relationship of chronic pain, mental illness and organic disorders. *Pain*, **32**, 185–95.
5. Benjamin, S. and Main, C.J. (1995). Psychiatric and psychological approaches to the treatment of chronic pain: concepts and individual treatments. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 188–213. Oxford University Press.
6. Asmundson, J.G., Coons, M.J., Taylor, S., *et al.* (2002). PTSD and the experience of pain; research and clinical implications of shared vulnerability and mutual maintenance models. *Canadian Journal of Psychiatry*, **47**, 930–7.
7. *Otis, J.D., Keane, T.M., and Kerns, R.D. (2003). An examination of the relationship between chronic pain and post-traumatic stress disorder. *Journal of Rehabilitation Research and Development*, **40**, 397–406.

8. British Pain Society. (2006). *Pain and substance misuse: improving the patient experience*. British Pain Society, London.
9. *Williams, A.C.deC. (2004). Assessing chronic pain and its impact. In *Psychosocial aspects of pain: a handbook for healthcare professionals, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
10. Fordyce, W.E. (1985). The behavioural management of chronic pain: a response to critics. *Pain*, **22**, 113–25.
11 Keefe, F.J. and Williams, D.A. (1992). Assessment of pain behaviours. In *Handbook of pain assessment* (eds. D.C. Turk and R. Melzack), pp. 277–92. Guilford Press, New York.
12 De Jong, J.R., Vlaeyen, J.W.S., Onghena, P., *et al.* (2005). Reduction of pain-related fear in complex regional pain syndrome type I: the application of graded exposure in vivo. *Pain*, **116**, 264–75.
13. *DeGood, D.E. and Shutty, M.S. (1992). Assessment of pain beliefs, coping and self-efficacy. In *Handbook of pain assessment* (eds. D.C. Turk and R. Melzack), pp. 214–34. Guilford Press, New York.
14. Engel, G. (1959). 'Psychogenic' pain and the pain prone patient. *The American Journal of Medicine*, **26**, 899–918.
15. Morley, S. (2004). What impact does childhood experience have on the childhood development of chronic pain? In *Psychosocial aspects of pain: a handbook for healthcare professionals, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
16. McBeth, J., Morris, S., Benjamin, S., *et al.* (2001). Associations between adverse events in childhood and chronic widespread pain in adulthood: are they explained by differential recall? *The Journal of Rheumatology*, **28**, 2305–9.
17. *Poleshuck, E.L. and Dworkin, R.H. (2004). Risk factors for chronic pain in patients with acute pain and their implications for prevention. In *Psychosocial aspects of pain: a handbook for healthcare providers, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
18. *Linton, S.J. (2004). Environment and learning factors in the development of chronic pain and disability. In *Psychological methods of pain control: basic science and clinical perspectives, progress in pain research and management*, Vol. 29 (eds. D.D. Price and E.M. Bushnell). IASP Press, Seattle.
19. McBeth, J., Macfarlane, G.J., Benjamin, S., *et al.* (2001). Features of somatization predict the onset of chronic widespread pain. Results of a large population based study. *Arthritis and Rheumatism*, **44**, 940–6.
20. *Gallagher, R.M. and Verma, S. (2004). Mood and anxiety disorders in chronic pain. In *Psychosocial aspects of pain: a handbook for healthcare professionals, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
21. Atkinson, J.H., Meyer, J.M., and Slater, M.A. (2004). Principles of psychopharmacology in pain treatment In *Psychosocial aspects of pain: a handbook for health care professionals, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
22. Sindrup, S., Otto, M., Finnerup, N., *et al.* (2005). Antidepressants in the treatment of neuropathic pain. *Basic and Clinical Pharmacology and Toxicology*, **96**, 399–409.
23. *Waters, S.J, Campbell, L.C., Keefe, F.J., *et al.* (2004). The essence of cognitive-behavioral pain management. In *Psychosocial aspects of pain: a handbook for healthcare professionals, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
24. Morley, S., Eccleston, C., and Williams, A. (1999). Systematic review and meta-analysis of randomised controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. *Pain*, **80**, 1–13.
25. Cohen, M.J.M. and Campbell, J.N. (eds.) (1996). *Pain treatment centers at a crossroads: a practical and conceptual reappraisal*. IASP Press, Seattle.
26. Thomsen, A.B., Sorensen, J., and Sjogren, P. (2002). Chronic non-malignant pain patients and health economic consequences. *European Journal of Pain*, **6**, 341–52.
27. Haldorsen, E.M.H., Grasdal, A.L., Skouen, J.S., *et al.* (2002). Is there a right treatment for a particular patient group? Comparison of ordinary treatment, light multidisciplinary treatment, and extensive multidisciplinary treatment for long-term sick-listed employees with musculoskeletal pain. *Pain*, **95**, 49–63.
28. Peters, L., Simon, E.P., Folen, R.A., *et al.* (2000). The COPE program: treatment efficacy and medical utilisation outcome of a chronic pain management program at a major military hospital. *Military Medicine*, **165**, 954–60.

5.2.7 Chronic fatigue syndrome

Michael Sharpe and Simon Wessely

Introduction

Chronic fatigue syndrome is a controversial condition, conflicts about which have frequently burst out of the medical literature into the popular media. Whilst these controversies may initially seem to be of limited interest to those who do not routinely treat such patients, they also exemplify important current issues in medicine. These issues include the nature of symptom-defined illness; patient power versus medical authority; and the uncomfortable but important issues of psychological iatrogenesis.[1,2] The subject is therefore of relevance to all doctors.

Fatigue as a symptom

Fatigue is a subjective feeling of weariness, lack of energy, and exhaustion. Approximately 20 per cent of the general population report significant and persistent fatigue, although relatively few of these people regard themselves as ill and only a small minority seek a medical opinion. Even so, fatigue is a common clinical presentation in primary care.[2]

Fatigue as an illness: chronic fatigue syndrome

When fatigue becomes chronic and associated with disability it is regarded as an illness. Such a syndrome has been recognized at least since the latter half of the last century. Whilst during the Victorian era patients who went to see doctors with this illness often received a diagnosis of neurasthenia, a condition ascribed to the effect of the stresses of modern life on the human nervous system the popularity of this diagnosis waned and by the mid-twentieth century it was rarely diagnosed (although the diagnosis subsequently became popular in the Far East—see Chapter 5.2.1). Although it is possible that the prevalence of chronic fatigue had waned in the population, it is more likely that patients who presented in this way were being given alternative diagnoses. These were mainly the new psychiatric syndromes of depression and anxiety, but also other labels indicating more direct physical explanations, such as chronic brucellosis, spontaneous hypoglycaemia, and latterly chronic Epstein–Barr virus infection.[2]

As well as these sporadic cases of fatiguing illness, epidemics of similar illnesses have been occasionally reported. One which

occurred among staff at the Royal Free Hospital, London in 1955 gave rise to the term myalgic encephalomyelitis (ME), although it should be emphasized that the nature and symptoms of that outbreak are dissimilar to the majority of those now presenting to general practitioners under the same label.

A group of virologists and immunologists proposed the term chronic fatigue syndrome in the late 1980s.[3] This new and aetiologically neutral term was chosen because it was increasingly recognized that many cases of fatigue were often not readily explained either by medical conditions such as Epstein–Barr virus infection or by obvious depression and anxiety disorders. Chronic fatigue syndrome has remained the most commonly used term by researchers. The issue of the name is still not completely resolved however: Neurasthenia remains in the ICD-10 psychiatric classification as a fatigue syndrome unexplained by depressive or anxiety disorder, whilst the equivalent in DSM-IV is undifferentiated somatoform disorder. Myalgic encephalomyelitis or (encephalopathy) is in the neurological section of ICD-10 and is used by some to imply that the illness is neurological as opposed to a psychiatric one. Unfortunately the case descriptions under these different labels make it clear that they all reflect similar symptomatic presentations, adding to confusion. Official UK documents have increasingly adopted the uneasy and probably ultimately unsatisfactory compromise term CFS/ME.[4] In this chapter, we will use the simple term chronic fatigue syndrome (CFS).

Clinical features

Symptoms

Chronic mental and physical fatigue, tiredness, or exhaustion that is typically exacerbated by activity is the core symptom of CFS. Commonly associated symptoms include impaired memory and concentration, muscular and joint pain, unrefreshing sleep, dizziness and breathlessness, headache, tender lymph glands, and sore throat. Patients often describe day-to-day fluctuations in symptoms, irrespective of activity. Periods of almost complete recovery may be followed by relapse, often described as sufficiently severe to make normal daily activity impossible. Depression and anxiety are common, and a proportion of patients suffer panic attacks.

Physical signs

Physical examination is typically unremarkable. Complaints of fever and lymphadenopathy are not confirmed on examination. The presence of definite physical signs (such as objectively measured fever) should not be ascribed to the syndrome and alternative diagnoses should be sought.

Other common characteristics

As well as the symptoms described above patients with CFS commonly have additional clinical characteristics. These are listed in Table 5.2.7.1.

Patients are often worried that remaining active despite fatigue will harm them and consequently avoid activity or oscillate between rest and bursts of activity, which produces fatigue, leading to a return to rest and so on.

Some patients feel strongly that their illness is 'medical' rather than 'psychiatric' and are particularly concerned that a psychiatric diagnosis implies that the illness is their fault, an indication of personal weakness or even an accusation of malingering. Perfectionist and high achieving lifestyles often with low underlying self-esteem are commonly observed in patients referred to hospital clinics.

Table 5.2.7.1 Common characteristics of patients with CFS

Thoughts beliefs and attitudes	Thought that symptoms indicate harm Belief that the illness is purely 'medical' Perfectionist attitudes
Coping behaviours	Avoidance of activities associated with symptoms Reduced activity level Oscillation in overall activity level
Physiology	Poor sleep Physiological deconditioning Effects of inactivity
Interpersonal and social	Dependence on carer Psychological iatrogenesis Occupational difficulties

Although there are no physical signs there may be measurable effects of reduced activity with so-called physiological deconditioning leading to poor tolerance of activity, and in cases where rest has been prolonged other physiological changes such as postural hypotension. Sleep is often unrefreshing and fragmented.

Some patients can become markedly dependent on a carer. Occupational stresses and difficulties are common and it can be difficult to determine if these were contributors to, or are consequence of their illness. Finally many patients have received unhelpful medical attention. Such psychological iatrogenesis includes, on the one hand dismissal of their complaints and on the other over investigation.[5]

Case study

A typical patient is found in the infectious disease department of the general hospital. She is a 30-years-old nurse and her principal complaints are of fatigue, poor concentration, and muscle pain. Her symptoms fluctuate and are made worse by physical and mental exertion. She is no longer able to work and has substantially reduced her daily activities. The history is of an acute onset of symptoms after a 'viral illness'. Enquiry reveals symptoms suggestive of depression or anxiety, but without obvious mood change. The patient strongly believes the illness to be 'medical' rather than 'psychiatric'.

Classification and diagnosis

There are several published case definitions for CFS. The currently most widely used definition is based on an international consensus of researchers is shown in Table 5.2.7.2.[6] A guide on its application has also been published.[7] It should be remembered that this definition represents nothing more than a working definition of a clinical problem, pending further understanding, and as with most psychiatric diagnoses, does not delineate a single disease.

Issues for a definition of chronic fatigue syndrome

The case definition shown in Table 5.2.7.2 has been useful in unifying the field and providing a widely used operational definition. However, it also has significant limitations.

Table 5.2.7.2 International consensus definition of chronic fatigue syndrome

1	Complaint of fatigue
	Of new onset
	Not relieved by rest
	Duration at least 6 months
2	At least four of the following additional symptoms
	Subjective memory impairment
	Sore throat
	Tender lymph nodes
	Muscle pain and joint pain
	Headache
	Unrefreshing sleep
	Post-exertional malaise lasting more than 24 h
3	Impairment of functioning
4	Other conditions that might explain fatigue excluded

(Reproduced from Fukuda, K. Straus, S.E. Hickie, I.B. *et al.* Chronic fatigue syndrome: a comprehensive approach to its definition and management, *Annals of Internal Medicine*, **121**, 953–9. Copyright 1994, The American College of Physicians.)

- It excludes fatigue associated with known organic disease.
- It overlaps with other functional medical diagnoses.
- It overlaps with psychiatric diagnosis.
- The homogeneity of the patient group it identifies is doubtful.

(a) Differentiation from fatigue associated with organic disease

Fatigue is a common symptom of most medical and psychiatric conditions. CFS refers only to fatigue where there is no clear alternative diagnosis (but does not exclude depression and anxiety unless the depression is of melancholic type or a manifestation of a bipolar disorder). It therefore only refers to idiopathic fatigue. This means that the definition highlights an important clinical problem but also means that the interesting equally important and probably informative phenomenon of fatigue in patients with diseases such as multiple sclerosis is excluded from this definition.

(b) Overlap with other medically unexplained syndromes

A number of medical diagnoses are defined only by symptoms. These functional syndromes are medical diagnoses where there is no identifiable pathology. They include chronic pain, fibromyalgia, and irritable bowel syndrome. Although chronic pain syndromes are principally characterized by pain, fibromyalgia by tender points, and irritable bowel syndrome by symptoms of bowel disturbance, all these syndromes are also associated with chronic fatigue, and patients diagnosed with one of these syndromes often meet the diagnostic criteria for CFS.[8]

(c) Overlap with psychiatric syndromes

Most patients who meet criteria for CFS also fulfil criteria for a psychiatric diagnosis. Many meet criteria for anxiety and depressive disorders and others merit diagnoses of somatoform disorder or neurasthenia. This issue is discussed further below.

(i) Depression

If patients with a depressive disorder are asked about a wide range of somatic symptoms including fatigue and/or muscle pain (which they are usually not) they often report these. If the diagnostic criteria for depressive disorders are applied to patients with fatigue a high proportion meet these.[9] Furthermore the prevalence of major depressive disorder in patients referred to hospital with CFS is substantially higher than in patients with chronic disabling medical diseases suggesting that depression is not simply a reaction to disability.[10] In practice, the diagnosis of depression can be difficult in patients presenting with fatigue: depressed mood is often not prominent and anhedonia can be hard to distinguish from the inability to pursue previously enjoyed activities because of fatigue. Finally, whilst there is a strong association between major depressive disorder and CFS, for as many as half of the patients seen in hospital clinics the symptoms cannot be readily given that diagnosis.

(ii) Anxiety disorders

Although less attention has been given to the association between fatigue and anxiety, an examination of diagnostic criteria for anxiety disorders reveals that the typical somatic symptoms of anxiety include fatigue and other symptoms listed as typical of CFS. If sought, generalized anxiety disorder can often be diagnosed in patients with CFS and panic can often be diagnosed in patients with severe episodic symptoms.[11] As with depression, however, anxious mood is rarely obvious and may be hard to distinguish from reasonable concern about consequences of being ill. Likewise, true phobic avoidance may be hard to distinguish from the consequences of fatigue and/or weakness.

(iii) Neurasthenia

ICD-10 differs from DSM-IV in including this diagnosis. It requires that the patient suffers from fatigue which is exacerbated by exertion, as well as several other somatic symptoms, and does not meet the criteria for a depressive or anxiety disorder (see Chapter 5.2.10). One study found that almost all of the referrals to a medical CFS clinic met the criteria for neurasthenia as defined by ICD-10.[12]

(iv) Somatoform disorders

According to DSM-IV patients with severe persistent fatigue who do not meet criteria for anxiety or depressive disorders are assigned to a somatoform disorder diagnosis. These are a controversial group of psychiatric syndromes characterized by medically unexplained symptoms and of presumed psychological origin.[13] There are a number of subcategories:

- Somatization disorder (Briquet's syndrome) is used to describe patients who report multiple, recurrent, medically unexplained symptoms; a minority of patients with CFS will meet the criteria for this disorder.
- Hypochondriasis describes a syndrome in which the patient's main concern is with the possibility that they are suffering from an organic disease. Whilst this diagnosis would seem to be applicable to many patients with CFS, it is problematic when the cause of the illness in question, which is regarded as uncertain by doctors as well as patients.
- Almost all patients with CFS not meeting the criteria for any of the above DSM disorders are likely to fall into the undemanding residual category in DSM-IV of 'undifferentiated somatoform disorder'. This diagnosis is of dubious practical use, and in effect merely confirms that the patient has multiple physical symptoms of unclear aetiology.

(v) Conclusion

Many patients with CFS meet the diagnostic criteria for a depressive or anxiety disorder, although in practice the presentation is often 'atypical'. It is likely that patients who do not meet the criteria for either of these could be diagnosed as having either neurasthenia (ICD-10) or undifferentiated somatoform disorder (DSM-IV).

Should we use the diagnosis of CFS?

From the psychiatrist's perspective it is parsimonious to ask whether a diagnosis of CFS is ever necessary or appropriate when the symptoms can always be described by a psychiatric diagnosis? This unsatisfactory situation is an artefact of parallel medical and psychiatric diagnostic systems for patients with somatic symptoms unexplained by disease. Consequently whether one uses a 'medical' diagnosis of CFS or a 'psychiatric' diagnosis of somatoform disorder is merely a matter of choice. When making that choice the following must be considered:

- A diagnosis of CFS only describes a presenting clinical syndrome, rather than a specific disorder or disease process.
- Pragmatically the relative acceptability of the alternative diagnosis to the patient is important. There is no point in giving a diagnosis that is rejected by the patient and impedes any therapeutic relationship and chances of treatment.
- One approach to overcoming the issue of parallel classification systems is to combine the medical diagnosis of CFS and the psychiatric diagnoses: According to such a scheme CFS would be subclassified into CFS/depression, CFS/anxiety, and CFS without depression or anxiety disorder (i.e. CFS/somatoform or CFS/neurasthenia). Psychiatric diagnoses that have important clinical utility such as major depressive disorder should obviously be made if present. The usefulness of diagnoses such as undifferentiated somatoform disorder is less clear.
- Finally rather than becoming side-tracked by the unanswerable question of whether the patients symptoms are ultimately 'medical' or 'psychiatric' in nature an open-minded and pragmatic approach is required.

Epidemiology

Prevalence

Fatigue is common but CFS is rare. As it can be difficult to differentiate CFS from depressive and anxiety disorders estimates that have attempted to exclude these diagnoses are lower than those that have not. Population studies in the United Kingdom and United States suggest that only approximately 0.5 per cent of the population can be regarded as having CFS.[1] Most of these persons are aged between 20 and 40 with a predominance of females. The syndrome is also seen in children and adolescents but less commonly.

Epidemics

Epidemics of a chronic fatigue-like syndrome have been described from various parts of the globe. This observation is compatible with, but does not establish, an infective cause. It remains unclear whether these were true epidemics and also whether the clinical picture reported is similar to that of cases of sporadic chronic fatigue syndrome.

Aetiology

Limitations in the available data

Although a considerable amount of research has been devoted to investigating the nature and causes of CFS there are few firm conclusions that can be drawn. This partly because many of the studies have had major methodological shortcomings:

- Patients have often been recruited from tertiary care clinics, using various diagnostic criteria inconsistently applied.
- Only a minority of studies have included comparison groups of patients with diagnoses of depression or anxiety disorders.
- Because most studies have used a case–control design, it is often impossible to know whether the findings they report are causes or consequences of the illness (for instance, as the result of reduced activity or sleep disturbance).
- CFS is almost certainly heterogeneous.

Considering these caveats there are a number of areas where positive findings have been reported.

Pathophysiology

Clinical observations of patients with CFS have led to the investigation of a number of hypotheses about the underlying pathophysiological mechanisms.

(a) Genetics

Twin studies suggest that CFS is moderately heritable.[1] Preliminary studies suggesting the involvement of specific genes require replication. Gene expression studies are likewise in their infancy—one problem being that the number of genes studied usually exceeds by several orders of magnitude the numbers of patients studied. Another problem is confounders—a large Swedish twin registry study for example suggested that genetic factors contributed to the risk of CFS both directly and via personality type.[14]

(b) Cardiovascular and respiratory abnormalities

Several investigators have reported abnormalities in the cardiorespiratory systems that may underpin the exercise intolerance. Hyperventilation has been suggested as a mechanism of symptom production, but only a minority of patients have biochemically confirmed hyperventilation. Low blood pressure has long been associated with the symptom of fatigue, and in some parts of Europe unexplained fatigue is confidently ascribed to this. Postural hypotension has been noted in some patients and whilst this may be a cause of fatigue it may also be a consequence of inactivity. Finally, various abnormalities in cardiac function have also been reported but are of uncertain significance.

(c) Infection

Perhaps because patients commonly describe their illness as beginning with 'flu-like symptoms' and because of the apparent epidemics, many investigators have sought objective evidence of an initiating or ongoing viral infection. Viral infection can probably initiate CFS. A prospective follow-up of people with positive evidence of acute Epstein–Barr infection did find that some patients went on to develop a fatigue syndrome.[15] Other infectious agents that may trigger CFS include Q fever and viral meningitis. If viruses do play a role in precipitating CFS, it would appear that it is only when certain types of viruses infect vulnerable persons. If CFS can be

precipitated by viral infection does persistence of the virus cause the ongoing symptoms? On current evidence the answer to this question seems to be no.

(d) Immune dysfunction

The evidence for an association between immunological abnormalities and CFS is more consistent than that for infective agents, with several studies suggesting abnormalities in lymphocyte numbers and function. However, similar changes can be found in patients with depressive disorders, and, although some studies have attempted to control for emotional disorder, both the specificity and causal importance of these observations remain unclear.[16]

(e) Sleep

Unrefreshing sleep is an almost ubiquitous complaint of people suffering from CFS. While studies have identified major sleep disorders such as sleep apnoea and narcolepsy as alternative diagnoses for small number of patients with daytime fatigue, simple disruption of slow-wave sleep is a much more common observation. While inefficient sleep could contribute to the daytime fatigue reported in both conditions, its specificity and aetiological role are uncertain.[16]

(f) Neuroendocrine and neurotransmitter abnormalities

The prominent fatigue of Addison's disease has led to the hypothesis that adrenal function is impaired in patients with CFS. In support of this suggestion there is reasonable evidence that patients with chronic fatigue and fibromyalgia have both low levels of cortisol, a point of difference from major depression. However it remains unclear if this apparent abnormality is cause or effect of the illness and associated inactivity. Patients with CFS also have evidence of abnormal functioning of cerebral serotonergic systems, which differ from those found in patients with depression. Like the abnormalities in adrenal function these findings are preliminary but of potential interest.[17]

(g) Brain imaging

Finally, a variety of techniques have been used to examine both the function and structure of the brain in patients with CFS. Cerebral perfusion studies have shown abnormalities, although similar, if not identical, abnormalities are also found in patients with depression. Possible white-matter changes reported on magnetic resonance scans are more controversial, and harder to interpret.[1,16]

(h) Conclusions

Despite a considerable research effort, so far no single pathophysiological process has been conclusively identified as causal of CFS. There is some evidence for a loss of physical fitness and possibly for abnormalities of neuroendocrine function. Viral infection can play a role as precipitating agent, although its importance as a perpetuating factor is less certain. Immunological abnormalities are common but of uncertain specificity, and appear not to be related to chronic symptoms. The current attention on neuroendocrine function takes the focus of investigation closer to those features known to be associated with depressive states. However, the evidence suggests that the changes in neurotransmitter and neuroendocrine function in patients with CFS may differ from those commonly observed in patients diagnosed with depressive disorder. Further studies are needed to confirm all these abnormalities and to clarify whether they are causal or merely epiphenomena.

Psychopathology

If there are no substantial biological abnormalities are there psychological ones? The initial psychiatric explanation of CFS was that it was misdiagnosed depressive disorder. However, whilst such misdiagnosis does occur, more complex explanations are required to adequately explain many cases of CFS.

(a) Somatization

It has been hypothesized that depression may still be 'behind' CFS even if not apparent. It is argued that a process referred to as 'somatization' (making the mental physical) is responsible. Whist the idea that the somatic symptoms of CFS are readily understandable as part of an emotional disturbance is a parsimonious alternative to some of the more elaborate mechanisms outlined above, as there is no 'marker' for somatization this hypothesis is hard to prove.

(b) Attribution

Patients attending specialist clinics with CFS typically attribute their illness to organic disease even when no evidence of this can be found by their physicians. Perhaps more importantly, some strongly resist psychological explanations for their illness, although most take a more mixed view. Whether these patients are biased in their views about illness or simply wiser than their physicians is unclear. However, strong and exclusively physical disease attributions may be a marker for an important illness-perpetuating process in CFS as they predict a poorer clinical outcome.

(c) Perceptual processes

Patients with CFS report a greater sense of effort in response to both psychological and physical demands than is explicable from objectively measured impairments. This observation raises the possibility that they are especially sensitive to bodily sensations such as effort, that is they 'amplify' or focus on them. Thus, one may hypothesize that as in panic disorder, the patients' beliefs about their symptoms may lead them to focus attention on to bodily sensations. Although plausible there is so far only limited evidence that this process is important in patients with CFS.[1]

(d) Coping behaviour

A tendency to avoid activities that exacerbate symptoms has been shown to occur in patients with CFS. The avoidance may be persistent or episodic in response to exacerbations of symptoms resulting in a 'boom and bust' pattern. Avoidance is associated with persistent disability, and has been suggested as the mechanism by which disease attributions for symptoms predicts poor outcome.[1]

(e) Personality characteristics

Both research studies and clinical experience suggest that many persons with CFS have a tendency towards hard driving, perfectionist, or obsessive–compulsive personalities, and associated overactive lifestyles. Evidence from the UK 1946 birth cohort for example indicates that ratings of physical activity in early life predict later CFS. In a large prospective Swedish twin registry measure of stress and emotionality are consistently associated with subsequent CFS. Those CFS sufferers may be predisposed to becoming physically and emotionally exhausted, and biased towards presenting emotional distress in a somatic form.[14]

(f) Stigma, misinformation, and iatrogenesis

Psychiatric diagnoses are stigmatized in the popular mind as indicating weakness or even unreal illness. Patients with CFS may be

susceptible to those social pressures and consequently prefer a medical diagnosis for their distress and inability to function. It has also been suggested that CFS may serve as culturally defined function of social communication, allowing a socially acceptable and hence 'non-psychiatric' expression of distress and protest about intolerable occupational and personal pressures. Much the same was said of neurasthenia in the People's Republic of China (see Chapter 5.2.1).

Another potentially important social factor is the controversy and the often misleading information about the illness that patients are exposed to. Self-help books the media and some doctors have frequently given the impression that the medial profession is more divided than it actually is in its understanding of CFS and have emphasized 'medical' explanations such as myalgic encephalopathy (ME) as the only appropriate diagnosis.

(g) Conclusions

Psychological and social factors are important perpetuating factors and include focusing of attention on symptoms avoidance of activity and a strong and exclusive medical disease attribution.

A comprehensive view of the cause of CFS

It now seems clear that rather than regarding pathophysiological and psychopathological studies as separate and competing approaches to the problem, it is more useful to consider a formulation of CFS that combines these factors. Table 5.2.7.3 summarizes relevant aetiological factors.

According to this integrated scheme causal factors are divided into those that may predispose to the illness, those that precipitate it, and those that perpetuate an established illness. Predisposing factors include previous episodes of major depressive disorder, and perhaps also certain personality characteristics, particularly achievement orientation and perfectionism associated with chronic stress, especially occupational stress. The precipitation of CFS by a viral infection is clinically plausible and proven in certain circumstances, whilst life stresses also seem to be important. Perpetuating factors may include neuroendocrine dysfunction, emotional disorder, and physical disease attributions, as well as coping by avoidance, chronic unresolved personal difficulties, and misinformation about the illness. They may be effectively combined in a cognitive–behavioural model of the illness that provides a basis for treatment with CBT.[1,18]

Course and prognosis of CFS

Anecdotal reports of the prognosis of CFS make gloomy reading. What is more, systematic studies are hardly more encouraging, suggesting that the commonest outcome of those attending a specialist CFS clinic is continuing ill health, up to and beyond 5 years.[19] However, these observations need elaboration. The rather dispiriting prognostic studies all refer to patients seen in specialist centres. Nearly all had several years of illness prior to referral, and it is unsurprising to find that chronicity predicts chronicity. Patients seen in specialist clinics often have strong views about illness aetiology and illness management that may negatively influence their acceptance of and adherence to potentially effective treatment. Primary care and community samples and patients appear to have a better outcome, as do children and adolescents. Finally, the current generation of outcome studies refer to the situation without treatment as potentially effective treatment was rarely given. Later sections of this chapter suggest that this view requires revision.

Table 5.2.7.3 An aetiological formulation of CFS

	Predisposing	Precipitating	Perpetuating
Biological	Genetics Previous depression	Infection	Effects of inactivity CNS dysfunction Reduced HPA activity
Psychological	Personality (perfectionism)	Response to stressor	Focus on symptoms Disease attribution Avoidant coping
Social	Chronic stress	Social/occupation stress	Life conflicts Psychological iatrogenesis

Evidence for treatments

Drug treatment

Many pharmacological treatments have been suggested for CFS. To date, none are of proven efficacy and several are potentially harmful.[4, 20] The evidence for antidepressant agents is mixed. Of available agents none is clearly superior for this patient group, although clinical experience suggests that the selective serotonin-reuptake inhibitor antidepressants may be better tolerated.

Graded activity (exercise) therapy

Well-conducted randomized controlled trials suggest that graded increases in physical activity is helpful in improving function and relieving symptoms.[4, 20]

Cognitive behaviour therapy

Systematic reviews have concluded that the strongest evidence for efficacy is for a rehabilitative type of cognitive–behaviour therapy (CBT).[4, 20]

Practical management

Assessment

Both a medical and psychiatric assessment is required in every case of suspected CFS.[21]

(a) Excluding organic disease

A small minority of those patients who present with severe chronic fatigue will be found to have occult organic disease. How frequently organic disease is found will depend on how thorough an assessment the patient has already received. The differential diagnosis is listed in Table 5.2.7.4.

There are no specific diagnostic tests and no characteristic abnormalities on laboratory investigations in CFS. Tests are conducted purely to exclude other diseases. All patients should have a full blood count, erythrocyte sedimentation rate or C-reactive

Table 5.2.7.4 Conditions to be considered in the differential diagnosis of chronic fatigue syndrome

Nature of symptoms	Possible condition
General	Occult malignancy Autoimmune disease Endocrine disease Cardiac, respiratory, or renal failure
Gastroenterological Neurological	Malabsorption including celiac disease Disseminated sclerosis Myasthenia gravis Parkinson's disease Early dementia Cerebrovascular disease
Infectious disease	Chronic active hepatitis (B or C) Lyme borreliosis HIV Tuberculosis
Respiratory disease	Nocturnal asthma Obstructive sleep apnoea
Chronic toxicity	Alcohol Solvents Heavy metals Irradiation
Psychiatric	Major depressive disorder Dysthymia Anxiety and panic disorder Somatoform disorder

protein, basic biochemistry screen, creatine kinase, random blood glucose, urine analysis, thyroid function, and possibly antinuclear antibody tests. Further investigation depends on the clinical findings and differential diagnoses under consideration. In our clinical experience unusual clinical features, such as weight loss, an absence of mental fatigue/fatigability, or a history of recent foreign travel, should all increases suspicion of alternative diagnoses.

(b) Excluding psychiatric diagnoses

All patients should have a psychiatric history taken and their mental state examined. The assessment should seek evidence of major depression, anxiety, and panic disorder, and also evaluate any suicidal intent. The psychiatric assessment should be systematic, as hidden distress is common and casual estimates of the patient's degree of distress may be misleading.

Making the diagnosis

As explained above the choice of diagnosis should be pragmatic; there is little merit in giving a diagnosis of CFS if the patient's symptoms are clearly those of depression or anxiety. In other cases, a diagnosis of CFS may be the most appropriate and useful; it offers the patient a coherent label for their symptoms and will therefore lessen the risk that they will embark on a fruitless search for a 'better' explanation. Above all, it is most important that neither the physician nor the patient stops at this diagnosis, but goes on to explain what it does and does not mean.

Making a formulation

An adequate individual patient assessment must identify all the important obstacles to recovery. It often needs to go beyond diagnosis to include a systematic individualized description of the aetiological factors in each case. These should include those factors listed in Table 5.2.7.1.

(a) Case formulation

A multidimensional description of the patient's illness provides a comprehensive picture of the factors that may be relevant to the patient's illness and is an important supplement to diagnosis. Its use can be illustrated by returning to the case example described above.

Case Study

Assessment of the patient described earlier revealed that she believed that her symptoms were by an ongoing virus infection and associated immune disturbance and that she should beware of exacerbating them. She consequently avoided activity and had been profoundly inactive for over a year, often lying in bed and sleeping for long periods. Therefore she had become physiologically deconditioned. She was frustrated with her inability to do things and sometimes felt low in mood about her predicament. Her previous job had been very stressful, but since becoming ill she had been unable to work. She had now lost her job and was cared for by her mother who also believed she had a permanent disability. Her doctor said that the best thing was rest. She had rejected a psychiatric consultation but was paying to see an alternative medicine therapist.

The findings can be summarized in an individualized form of Table 5.2.7.3 with an emphasis on the perpetuating factors, which can be seen as reversible barriers to recovery.

General management

The five basic steps essential to the care of patients with CFS are listed in Table 5.2.7.5.

The doctor should listen to the patient's story and ask about his or her own understanding of the illness. It is usually also worth seeing the partner or relevant family members. It is important to address misunderstandings about the nature of the illness and especially to make clear that it is not progressive or life threatening. A positive explanation of CFS as a 'dysfunction of the central nervous system' emphasizing reversibility is often helpful. Anxiety and depression can be explained as understandable consequences of illness and treatment given for them. The adverse physiological and psychological effects of prolonged bed rest should be explained, and the patient encouraged to avoid extremes of both inactivity

Table 5.2.7.5 Principles of management of CFS

1. Acknowledge the reality of the patient's symptoms and disability
2. Provide appropriate education about the nature of the syndrome
3. Treat identifiable depression and anxiety disorder
4. Encourage a very gradual return to normal functioning
5. Help the patient overcome occupational and interpersonal obstacles to normal functioning

and exertion. Finally there are often problem in relationships and with employers than the patient may need help to address.

An initial hospital appointment that achieves all the above usually requires at least 45 min. An evidence-based self-help book may be recommended.[22]

(a) Pharmacological treatments

Patients are often reluctant to take antidepressants and careful explanation and follow-up are required. Other pharmacological agents should only be used with care and preferably only as part of randomized controlled trials.

(b) Graded activity and exercise therapy

This should be considered for patients who are physically inactive. However, they need to be slowly graded and tailored to the patients' ability and progress; a simplistic application of fixed exercise regimens, particularly if given without adequate explanation is unlikely to be helpful, and may exacerbate symptoms and damage confidence.

(c) CBT

Whilst CBT is currently the mainstay of management it is not effective in all patients and requires both careful explanation and therapists skilled in its delivery to patients with CFS. It is particularly important that patients do not interpret referral for a behavioural treatment as the doctor implying that their illness is imaginary. It can be explained that the most likely cause of fatigue in CFS is changes in brain and neuroendocrine function and that these can be reversed by these therapies. It is also both useful and true to draw attention to studies showing the effective of CBT in improving symptoms, quality of life and outcome in conditions such as diabetes, cancer, and rheumatoid arthritis. General physical rehabilitation services may be useful for patients with chronic severe disability.

Potential management problems

Several issues may complicate the management of patients with CFS. These include:

(a) Strong illness beliefs

Difficulties are most likely to arise when patient and physician hold differing beliefs about the nature and best management of the illness. This problem can often be overcome by the physician acknowledging the patient's beliefs without either agreeing with them or arguing about them. If the patient's family, friends, or acquaintances suggest or encourage views that the physician regards as unhelpful the problem is more difficult and a meeting with the other parties may be necessary.

(b) Alternative therapies

Patients with CFS often turn to alternative and complementary medicine. Whilst some of these therapies may be easily continued in parallel with rehabilitative management, others may interfere either by the explanation of the illness they offer or by their practical requirements. In such cases it can be helpful to explain the need to pursue one approach at a time in order to learn what helps.

(c) Official reports and financial benefits

Perhaps the greatest difficulty is when patients ask the physician to write reports on their behalf, confirming that they suffer from permanent disability. The physician wants to help the patient and ensure that they get appropriate benefits but also to avoid a self-fulfilling prophecy. This dilemma has no easy solution, but it seems important not to confirm a negative prognosis until potentially effective treatment has been tried.

(d) Poor prognosis patients

For patients who have been identified as having a poor prognosis because of a long history of severely impaired functioning, or poor response to treatment, regular (albeit infrequent) long-term follow-up is, at the least, likely to limit iatrogenic harm from unnecessary investigations and ineffective treatments. Often in such cases the best strategy is simply to tell the patient how well they are managing in difficult circumstances.

Possibilities for prevention

We do not know how to prevent CFS, but its development can probably be modified.

The most important place for such intervention may be the transition from an acute fatigue state to chronic disability. For example, although most of us have been exposed to Epstein–Barr virus infection by the time we reach 30 years of age, few go on to develop CFS. Encouraging modest amounts of activity in the weeks after an acute infection has been shown to be effective in reducing the duration of symptoms. Intervention that maintain activity and prevent a slide into a vicious circle of symptoms, reduced activity, demoralization, disability, and depression might therefore offer an opportunity for prevention. The second area for intervention is in achieving better attitudes to symptoms and distress. Simplistic depictions of illness as either physical or psychological and the corresponding division in medical services are clearly unhelpful. The third area is in employment practice. CFS is often associated with work stress and dissatisfaction and only a minority of employers allow a flexible return to work. Finally, a doctor–patient relationship that both allows the patient to be ill and encourages recovery is probably the most important preventive strategy.

Conclusions and future directions

Chronic fatigue syndrome is best regarded as a descriptive term for a clinical presentation, rather than as a discrete condition. The group of patients it defines is almost certainly aetiologically heterogeneous. While psychiatric diagnosis provides one approach to subclassification of CFS, both the medical and psychiatric current diagnostic systems have significant limitations, and a multidimensional description of the patient's characteristics may be more clinically useful.[23]

The illness defined by the term chronic fatigue syndrome is important because it represents potentially treatable disability and suffering. It is also important because the clinical problems it gives rise to highlight shortcomings in our present approach to illness. Whatever is ultimately discovered about the causes of CFS, the attention it is receiving offers a golden opportunity to reappraise our understanding and classification of human illness and to re-examine our current organization of medical care.

Further information

Wessely, S., Sharpe, M., and Hotopf, M. (1998). *Chronic fatigue and its syndromes*. Oxford University Press, Oxford.

National Institute for Health and Clinical Excellence Guideline number 53. Chronic fatigue syndrome/ myalgic encephalomyelitis (or encephalopathy); diagnosis and management. http://guidance.nice.org.uk/CG53

References

1. Prins, J.B., Van der Meer, J.W., and Bleijenberg, G. (2006). Chronic fatigue syndrome. *Lancet*, **367**, 346–55.
2. Wessely, S., Hotopf, M.H., and Sharpe, M. (1998). *Chronic fatigue and its syndromes*. Oxford University Press, Oxford.
3. Holmes, G.P., Kaplan, J.E., Gantz, N.M., *et al.* (1988). Chronic fatigue syndrome: a working case definition. *Annals of Internal Medicine*, **108**, 387–9.
4. Baker, R. and Shaw, E.J. (2007). Diagnosis and management of chronic fatigue syndrome or myalgic encephalomyelitis (or encephalopathy): summary of NICE guidance. *British Medical Journal*, **335**, 446–8.
5. Sharpe, M. (1998). Doctors' diagnoses and patients' perceptions: lessons from chronic fatigue syndrome. *General Hospital Psychiatry*, **20**, 335–8.
6. Fukuda, K., Straus, S.E., Hickie, I.B., *et al.* (1994). Chronic fatigue syndrome: a comprehensive approach to its definition and management. *Annals of Internal Medicine*, **121**, 953–9.
7. Reeves, W.C., Lloyd, A., Vernon, S.D., *et al.* (2003). Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. *BMC Health Services Research*, **3**, 25.
8. Wessely, S., Nimnuan, C., and Sharpe, M. (1999). Functional somatic syndromes: one or many? *Lancet*, **354**, 936–9.
9. Skapinakis, P., Lewis, G., and Meltzer, H. (2000). Clarifying the relationship between unexplained chronic fatigue and psychiatric morbidity: results from a community survey in Great Britain. *The American Journal of Psychiatry*, **157**, 1492–8.
10. Katon, W., Buchwald, D.S., Simon, G.E., *et al.* (1991). Psychiatric illness in patients with chronic fatigue and rheumatoid arthritis. *Journal of General Internal Medicine*, **6**, 277–85.
11. Fischler, B., Cluydts, R., De Gucht, Y., *et al.* (1997). Generalized anxiety disorder in chronic fatigue syndrome. *Acta Psychiatrica Scandinavica*, **95**, 405–13.
12. Farmer, A., Jones, I., Hillier, J., *et al.* (1995). Neurasthenia revisited. *British Journal of Psychiatry*, **167**, 496–502.
13. Mayou, R., Kirmayer, L.J., Simon, G., *et al.* (2005). Somatoform disorders: time for a new approach in DSM-V. *American Journal of Psychiatry*, **162**, 847–55.
14. Kato, K., Sullivan, P.F., Evengard, B., *et al.* (2006). Premorbid predictors of chronic fatigue. *Archives of General Psychiatry*, **63**, 1267–72.
15. White, P.D., Thomas, J.M., Amess, J., *et al.* (1995). The existence of a fatigue syndrome after glandular fever. *Psychological Medicine*, **25**, 907–16.
16. Afari, N. and Buchwald, D. (2003). Chronic fatigue syndrome: a review. *American Journal of Psychiatry*, **160**, 221–36.
17. Cho, H.J., Skowera, A., Cleare, A., *et al.* (2006). Chronic fatigue syndrome: an update focusing on phenomenology and pathophysiology. *Current Opinion in Psychiatry*, **19**, 67–73.
18. Surawy, C., Hackmann, A., Hawton, K.E., *et al.* (1995). Chronic fatigue syndrome: a cognitive approach. *Behaviour Research and Therapy*, **33**, 535–44.
19. Cairns, R. and Hotopf, M. (2005). A systematic review describing the prognosis of chronic fatigue syndrome. *Occupational Medicine (London)*, **55**, 20–31.
20. Chambers, D., Bagnall, A.M., Hempel, S., *et al.* (2006). Interventions for the treatment, management and rehabilitation of patients with chronic fatigue syndrome/myalgic encephalomyelitis: an updated systematic review. *Journal of the Royal Society of Medicine*, **99**, 506–20.
21. Sharpe, M., Chalder, T., Palmer, I., *et al.* (1997). Chronic fatigue syndrome. A practical guide to assessment and management. *General Hospital Psychiatry*, **19**, 185–99.
22. Campling, F. and Sharpe, M. (2000). *Chronic fatigue syndrome: the facts*. Oxford University Press, Oxford.
23. Sharpe, M., Mayou, R., and Walker, J. (2006). Bodily symptoms: new approaches to classification. *Journal of Psychosomatic Research*, **60**, 353–6.

5.2.8 Body dysmorphic disorder

Katharine A. Phillips

The dysmorphophobic patient is really miserable; in the middle of his daily routines, talks, while reading, during meals, everywhere and at any time, he is caught by the doubt of deformity . . . . Enrico Morselli, 1891[1]

Introduction

Body dysmorphic disorder (BDD), also known as dysmorphophobia, is a relatively common, severe, and sometimes difficult-to-treat condition that has been described for more than a century.[1–3] BDD consists of a distressing or impairing preoccupation with an imagined or slight defect in one's physical appearance. BDD is classified as a separate disorder in DSM-IV and a type of hypochondriasis in ICD-10. This disorder can cause severe distress and notably impaired functioning. In addition, risk behaviours—suicidality, violence, problematic substance use, and compulsive tanning—appear common in BDD. Despite its severity, BDD is underrecognized in clinical settings.

Clinical features

Demographic characteristics

BDD occurs in all age groups.[3–5] It appears about equally common in females and males or may be somewhat more common in females.[3] Most individuals with BDD have never been married, and a high proportion is unemployed, often because of their psychopathology.[3–7]

Bodily preoccupations

People with BDD are preoccupied with the idea that some aspect of their appearance is ugly, unattractive, deformed, flawed, or defective in some way.[1–6, 8–10] Concerns usually focus on the face or head but can involve any body area.[3–6, 8–10] Skin (e.g. acne, scars, lines, or pale skin), hair (e.g. thinning or excessive body or facial hair), and nose (e.g. size or shape) concerns are most common. Most patients are preoccupied with several body areas. The preoccupation usually focuses on specific areas but may involve overall appearance.

BDD preoccupations are distressing, time consuming (occurring for an average of 3–8 h a day), and usually difficult to resist or control.[3] They are often associated with low self-esteem, shame, rejection sensitivity, and high levels of neuroticism, introversion,

depressed mood, anxiety, anger-hostility, and perceived stress.[3] Patients often believe that they are unacceptable—e.g. worthless, inadequate, unlovable, and an object of ridicule and rejection.[3,9]

Insight/delusionality

Insight is usually poor or absent; 27–39 per cent of patients are currently delusional (completely convinced that their belief is accurate and undistorted).[3,11,12] Most do not recognize that their belief is due to a mental illness or has a psychological/psychiatric cause.[3,12] In addition, a majority have ideas or delusions of reference, believing that others take special notice of the supposed appearance defects—for example, stare at them or mock the person because of how they look.[3,12] Referential thinking can fuel feelings of anger and rejection as well as social isolation.

Compulsive and safety behaviours

Nearly all patients perform BDD-related compulsive or safety behaviours (Table 5.2.8.1), which are time consuming (occurring for hours a day) and difficult to resist or control.[3–6,10] The behaviours usually aim to examine, improve, hide, or obtain reassurance about the perceived defects. These behaviours typically do not alleviate distress and may even worsen it.

Compulsive skin picking, which 27–45 per cent of BDD patients do to try and improve their appearance, can cause considerable skin damage.[3,5] Emergency surgery is sometimes required—for example when sharp implements used for picking rupture major blood vessels. Compulsive tanning to darken 'pale' skin or minimize perceived acne, scarring, or 'marks' can cause skin damage and may increase cancer risk.[3]

Psychosocial functioning and quality of life

Functioning and quality of life are usually very poor.[3,4,7] Some people, with effort, function adequately despite their distress, although usually below their potential. Those with severe BDD may be profoundly impaired by their symptoms—for example, housebound for years, unemployed and socially isolated, and chronically suicidal.

Social impairment is nearly universal.[3,7] People with BDD feel embarrassed and ashamed of their 'ugliness', are anxious around others as a result, and fear being rejected because of how they look. Thus, they may have few or no friends; avoid dating, physical intimacy, and other social interactions; or get divorced. Impairment in academic or occupational functioning is common, due to the time consuming and distracting nature of BDD symptoms and a desire to avoid interactions with others.[3,7] In a broadly ascertained BDD sample (n = 200), 36 per cent of individuals were not currently working and 32 per cent were not able to be in school or do school work because of psychopathology (BDD was the primary diagnosis for most).[7] In two BDD series, more than a quarter of individuals had been completely housebound for at least 1 week because of BDD symptoms, and more than 40 per cent had been psychiatrically hospitalized.[5,13] Mental health related quality of life is markedly poorer than for the general population and even poorer than for patients with diabetes, a recent myocardial infarction, or clinical depression.[7]

Table 5.2.8.1 Common compulsive and safety behaviors in body dysmorphic disorder

- Comparing one's appearance 'defects' with the same body areas of other people
- Checking the perceived defects directly, in mirrors, or in other reflecting surfaces
- Excessive grooming—for example, applying make-up, shaving, hair cutting or removal, hair styling
- Camouflaging—for example with a hat, clothes, sunglasses, hair, body position, hand, or make-up
- Seeking reassurance from others or trying to convince others of the 'defect's' ugliness
- Skin picking
- Excessive exercising or weightlifting
- Tanning
- Frequent clothes changing
- Touching the perceived defect
- Dieting
- Body measuring
- Compulsive shopping for beauty products, remedies, or clothes
- Seeking surgical, dermatologic, and other cosmetic treatment for the perceived deformity

Suicidality

Suicidal ideation and attempts appear very common. Reported lifetime rates of suicidal ideation and suicide attempts are 78–81 and 24–28 per cent, respectively.[9,13,14] Among adolescent inpatients, those with BDD have significantly greater suicidality than those without BDD. The rate of completed suicide, while preliminary, appears markedly high. In a prospective study, the annual suicide rate was 0.35 per cent, which is approximately 45 times higher than for the US population (adjusted for age, gender, and geographic region) and higher than for most other mental disorders.[15] A study of dermatology patients who committed suicide found that most had acne or BDD.[16] Indeed, individuals with BDD have many suicide risk factors.[3,14]

Aggression and violence

In several BDD studies, 36–38 per cent of patients reported lifetime aggression/violence due specifically to BDD symptoms.[3,10] Such behaviour may be fuelled by anger about looking 'deformed', an inability to fix the perceived defect, and misperceptions of being rejected, ridiculed, or mocked because of the appearance 'defects'. Individuals with BDD tend to misinterpret self-referent facial expressions as contemptuous and angry,[17] misinterpret ambiguous social (and other) situations as threatening,[18] and have high levels of anger/hostility.[3] Surgeons and dermatologists may be victims of violence—even murder—fuelled by dissatisfaction with the outcome of cosmetic procedures.[3] In a survey of 265 plastic surgeons, 12 per cent reported that a BDD patient had physically threatened them.[19]

Comorbidity

Major depressive disorder is the most frequently comorbid disorder, occurring in about 75 per cent of individuals with BDD.[5,13] Social phobia, OCD, and substance use disorders are also common.[5,8,13] Of note, one study found that 49 per cent of 200 BDD subjects had a lifetime substance use disorder, 70 per cent of whom reported that BDD contributed to their substance use.[5] Muscle dysmorphia,

a preoccupation with the idea that one's body is insufficiently lean or muscular, may lead to anabolic steroid abuse.[3]

Gender

Men and women appear to have largely similar clinical features.[10, 13, 20] However, in two United States studies (n = 200 and n = 188[13, 20]) males were more likely to be single, have a substance use disorder, and be preoccupied with thinning hair and small body build. Females were more likely to be preoccupied with weight, hips, and excessive body hair, and were more likely to pick their skin and use their hands or make-up for camouflage. One of these two studies, and a study from Italy (n = 58[10]), found that females were more likely to be preoccupied with their breasts/chest and legs, check mirrors, and use camouflaging. In all three studies, a concern with genitals was more common in males, and a comorbid eating disorder more common in females.

BDD in children and adolescents

BDD usually begins during early adolescence and can occur in childhood.[4, 5, 10] While data are limited, BDD's clinical features in youth appear largely similar to those in adults. Of note, children and adolescents appear to have lifetime rates of functional impairment similar to those in adults, despite having had fewer years over which to have developed these problems.[3] In the one study that directly compared adolescents to adults, adolescents were more likely to have delusional BDD beliefs and had a significantly higher lifetime suicide attempt rate (44 per cent versus 24 per cent), underscoring the importance of recognizing BDD in this age group.

Cross-cultural aspects of BDD

Case reports and series from around the world suggest that BDD's clinical features are generally similar across cultures but that cultural factors may produce nuances and accents on a basically invariant, or universal, expression of BDD.[3] It is unclear whether koro (a belief that one's penis is shrinking, which occurs primarily in Southeast Asia) is a form of BDD.

Classification and relationship to other disorders

A clinically important classification controversy is whether delusional and non-delusional BDD are the same or different disorders. Whereas BDD is classified as a somatoform disorder, its delusional variant is classified as a psychotic disorder—a type of delusional disorder. DSM-IV, however, allows delusional patients to be diagnosed with both BDD and delusional disorder, reflecting data suggesting that its delusional and non-delusional variants may constitute the same disorder, which spans a spectrum of insight. Indeed, delusional and non-delusional BDD appear more similar than different, although delusional BDD appears more severe.[11,12] Of clinical significance, delusional and non-delusional patients both appear to respond to serotonin-reuptake inhibitors (SRIs) as monotherapy but not to antipsychotic monotherapy.[21]

Another important question is whether BDD is related to OCD, social phobia, major depressive disorder, or eating disorders.[3] BDD is widely conceptualized as an OCD-spectrum disorder. Supporting this conceptualization, OCD often co-occurs with BDD, and BDD appears more common in first-degree relatives of OCD probands than control probands.[3] Data from a variety of domains suggest that BDD and OCD have many similarities.[3] However, BDD and OCD have some differences and do not appear to be identical disorders.[3] Although BDD's relationship to other disorders has received less investigation, preliminary data suggest that BDD may also be related to major depressive disorder. However, BDD does not appear to simply be a symptom of depression.[3] Although ICD-10 classifies BDD as a type of hypochondriasis, no studies have examined their relationship.

Diagnosis

BDD can be diagnosed using questions at the top of Table 5.2.8.2, which follow DSM-IV's diagnostic criteria. Clinicians should adequately probe for examples of clinically significant distress and impairment in social, occupational, and other aspects of functioning. BDD is diagnosed if the person is excessively preoccupied with a non-existent or slight physical flaw (for example, thinks about it for at least an hour a day), and the concern causes clinically significant distress or impairment in functioning. The appearance concerns should not be better accounted for by an eating disorder. However, BDD and eating disorders may co-occur, in which case both disorders should be diagnosed.

The bottom of Table 5.2.8.2 includes questions that are ***not*** recommended for screening for or diagnosing BDD. The word 'imagined' is problematic, because most patients have poor or absent insight and do not think their appearance problem is imagined. Terms such as 'deformed' or 'disfigured' are too extreme for some patients to endorse. Asking if there is something wrong with one's body is too broad, as patients may interpret this to refer to bodily functioning.

ICD-10's criteria require that patients persistently refuse to accept the advice and reassurance of several different doctors that they do not have an abnormality. However, many people with BDD do not disclose their appearance concerns to doctors or even seek medical care because they are housebound, ashamed, believe they cannot be helped, lack medical insurance, or do not have access to

Table 5.2.8.2 Questions to diagnose BDD

Recommended questions for diagnosing BDD:

1 Are you very worried about your appearance in any way? ***OR:*** Are you unhappy with how you look? ***If yes:*** Can you tell me more about your concern?

2 Does this concern preoccupy you? Do you think about it a lot and wish you could worry about it less? How much time would you estimate you spend each day thinking about how you look, if you were to add up all the time you spend?

3 What effect does this preoccupation with your appearance have on your life? Has it caused you a lot of distress (for example, anxiety or depression)? Has it significantly interfered with your social life, relationships, school work, job, or any other activities? Has it affected your family or friends?

*Questions that are **not** recommended when screening for or diagnosing BDD:*

1 It is recommended that patients ***not*** be asked if they are concerned with an 'imagined' defect in their appearance, whether they think they are 'deformed' or 'disfigured', or whether there is something 'wrong with (their) body'

2 To diagnose BDD, it should not be required that patients refuse to accept the advice and reassurance of doctors that they do not have an abnormality

health care for other reasons. Using this diagnostic criterion will underdiagnose BDD. BDD is also underidentified in many mental health databases because its ICD-9 diagnostic code identifies it as hypochondriasis.

BDD usually goes undiagnosed in clinical settings.[3,4,22] Sufferers often conceal their symptoms due to embarrassment and shame.[3,4,22] They may volunteer only depression, anxiety, or discomfort in social situations. The compulsive and safety behaviours in Table 5.2.8.1 may be clues to BDD's presence. BDD may be misdiagnosed as another disorder[3]: social phobia or agoraphobia (due to secondary social anxiety and isolation), panic disorder (because panic attacks may occur after looking in the mirror or when feeling scrutinized by others), trichotillomania (when hair is cut or plucked to improve perceived flaws, such as uneven eyebrows), or OCD (due to obsessional preoccupations and compulsive behaviours). Delusional patients are sometimes misdiagnosed with schizophrenia, psychotic depression, or psychotic disorder NOS. To diagnose BDD, patients must usually be asked directly about BDD symptoms.

Epidemiology

BDD has been reported in 0.7–1.7 per cent of community samples.[3] In the largest study, a nationwide survey in Germany (n = 2552), BDD was present in 1.9 per cent of women and 1.4 per cent of men.[23] In smaller non-clinical student samples, BDD's prevalence has ranged from 2.3–13 per cent.[3] A prevalence of 9–12 per cent has been reported in dermatology settings, 3–15 per cent in cosmetic surgery or plastic surgery settings, 8–37 per cent in patients with OCD, 11–13 per cent in social phobia, 26 per cent in trichotillomania, and 14–42 per cent in atypical major depression.[3] In a study of 122 general psychiatric inpatients, 13 per cent had BDD's, which was higher than for schizophrenia, OCD, PTSD, and eating disorders.[22]

Pathogenesis

BDD's pathogenesis is likely to be complex and multifactorial.[3] Aetiologic factors likely involve a complex interplay of genetic and environmental risk factors.[3] Preliminary data indicate an association of the $GABA_A$-γ2 gene with BDD. Environmental risk factors may include perceived childhood neglect and/or abuse, teasing, and low parental warmth. A role is also likely for socio-cultural and evolutionary pressures (e.g. symmetrical features may signal reproductive health).

Neuropsychological studies indicate a tendency to focus on isolated details of visual and verbal stimuli rather than more global, configurational attributes[24]—consistent with clinical observations that patients selectively attend to specific aspects of their appearance or minor flaws. Cognitive processing studies indicate that BDD patients tend to misinterpret ambiguous social (and other) situations as threatening and misinterpret self-referent facial expressions as contemptuous and angry.[17,18] These interpretive biases may combine with rejection sensitivity, perfectionism, and a focus on aesthetics to contribute to BDD's development.[3] High neuroticism and low extroversion may also play a role.[3] Many potential risk factors (e.g. neuroticism) are not specific to BDD, but the overall combination of risk factors may be. BDD's neurocircuitry is unknown but likely involves a complex interplay of dysfunction in several neural systems,[3,25] including circuitry involved in OCD (orbitofrontal cortex, anterior cingulate cortex, caudate, thalamus). BDD's shared features with other anxiety disorders and depression points to possible dysfunction in amygdala, prefrontal cortex, and anterior cingulate cortex. Brain regions involved in body image and facial emotion perception (e.g. right parietal cortex, amygdala, occipitotemporal cortex [e.g. fusiform face and extrastriate body areas]) may also be involved.

Course and prognosis

Prospective and retrospective studies indicate that BDD is usually chronic.[5,13,26] More severe BDD, a longer duration of BDD, and the presence of a personality disorder predict a lower probability of remission. However, when BDD is accurately identified and its treatment optimized, the prognosis appears much more favourable.

Treatment

BDD's treatment is described in more detail elsewhere, including in a guideline from the United Kingdom's National Institute of Clinical Excellence (NICE).[3,21,27,28] Serotonin-reuptake inhibitors (SRIs, or SSRIs) and cognitive–behavioural therapy (CBT) are currently recommended as the first-line treatments.[3,21,27,29] Treatment studies are limited, and more research is needed. However, available data consistently indicate that a majority of patients improve with these treatments.

Evaluation of treatments

Surgical, dermatologic, and other cosmetic treatment

A majority of BDD patients seek often-costly cosmetic treatment.[3,8,9,13,30] Dermatologists and surgeons are most often consulted, but any type of physician may be seen. It appears that most BDD patients are dissatisfied with such treatment.[3,9,13,30] Occasionally, dissatisfied patients sue, or are violent towards, the physician. Some patients perform their own surgery[3,6]—e.g. cutting open one's nose with a razor blade and trying to replace nose cartilage with chicken cartilage in the desired shape.

Pharmacotherapy and other somatic treatments

All SRI studies to date indicate that SRIs are often efficacious for BDD.[3,21] These studies include a placebo-controlled fluoxetine study (n = 67), a controlled and blinded cross-over study comparing the SRI clomipramine to the non-SRI antidepressant desipramine (n = 29), and open-label trials of fluvoxamine, citalopram, and escitalopram (n = 15–30). In these studies, 53 per cent to 73 per cent of patients responded to the SRI. The cross-over trial found greater efficacy for clomipramine than desipramine, suggesting that SRIs may be more efficacious than non-SRI antidepressants for BDD. This important finding is consistent with clinical series and retrospective data suggesting that SRIs are more efficacious than a broad range of non-SRI medications for BDD.[3,21]

Response to an SRI usually develops gradually and may require up to 12–14 weeks of treatment (while reaching a relatively high dose) to be evident.[3,21] Although dose-finding studies are lacking, relatively high SRI doses (higher than typically used for depression)

appear to often be needed.[3,21] Response to medication usually includes a decrease in appearance preoccupations, distress, and compulsive/safety behaviours, as well as improved functioning. Suicidality, depressive symptoms, anxiety, and anger-hostility often improve.[3] Of note, delusional patients often improve with SRI monotherapy, whereas limited data suggest that antipsychotic monotherapy is usually ineffective for delusional BDD.[3,21]

Patients who do not improve with one SRI may improve with another SRI.[3, 21] Regarding SRI augmentation, a small double-blind randomized controlled trial found that pimozide was not more efficacious than placebo as a fluoxetine augmenter.[3,21] Clinical series suggest that augmentation of an SSRI with buspirone, lithium, or clomipramine may be helpful.[3,21] (SSRIs may increase clomipramine blood levels, however, which may cause toxicity; thus, if this approach is tried, clomipramine should be started at a very low dose with monitoring of levels.) Clinical observations suggest that SSRI augmentation with venlafaxine, atypical neuroleptics, or bupropion may be helpful for some patients.[3,21]

Monotherapy with agents other than SRIs has not been well studied.[3, 21] A small open-label trial (n = 11) suggests that venlafaxine may be efficacious.[3] For severe and treatment-refractory cases an MAO inhibitor may be worth trying (but should never be combined with an SRI). Available case series and reports, while very limited, suggest that ECT is generally ineffective for BDD and secondary depressive symptoms.[3,21]

Clinical experience suggests that many patients relapse after SRI discontinuation and that long-term treatment may be needed, with efficacy usually sustained over time.[3,21] For patients who appear at high risk of suicide or violence, lifelong treatment with an effective SRI is recommended, as suicides have been known to occur after SRI discontinuation.

Cognitive behavioural therapy (CBT)

Preliminary data suggest that CBT is often efficacious for BDD.[3,27,29] CBT typically includes:[3,27–29]

1 *Cognitive restructuring* to identify cognitive errors and develop more accurate and helpful BDD-related thoughts and beliefs
2 *Behavioural experiments* to test the accuracy of BDD beliefs
3 *Exposure* to avoided situations (e.g. leaving the house, attending social gatherings)
4 *Response prevention* to decrease or stop compulsive behaviours (e.g. stopping excessive mirror checking, limiting grooming time)

In two randomized studies (n = 54 and 19), patients improved more with CBT than on a waiting list.[3,29] Several case series (n = 5–13) found that CBT was efficacious for BDD.[3,29] The number of sessions in these reports ranged from twelve 60-min sessions to sixty 90-min sessions.

Data on exposure and response prevention alone—without a cognitive element—is limited to a retrospective study and small case series with up to 10 subjects.[3,29] These reports note favourable outcomes, although in the author's experience cognitive approaches are a helpful, even necessary, component of treatment for most patients. It can be particularly helpful to work on core beliefs, which typically involve feelings of inadequacy and being unwanted by others.[3, 9] Clinical experience additionally suggests that the following are beneficial components of CBT:[3, 28]

1 *Mirror retraining*, which involves learning to see one's entire body in a non-judgemental and 'holistic' way (rather than focusing on disliked areas), while refraining from excessive mirror checking
2 *Habit reversal* for skin picking and repetitive hair pulling or plucking
3 *Mindfulness skills*
4 *Activity scheduling* and *scheduling pleasant activities* for more severely ill, depressed, and inactive patients
5 *Motivational interviewing*, which may be needed to engage and keep patients in treatment

Future studies (e.g. dismantling studies) are needed to determine which specific components of CBT are necessary and effective.

Other types of psychotherapy have not been well studied and are not currently recommended as first-line treatments.[3] Nonetheless, clinical experience suggests that insight-oriented or supportive therapy—in addition to an SRI and/or CBT—may help some patients cope with their illness or with co-occurring problems or disorders.[3]

Management

Management approaches are described in more detail elsewhere.[3, 21, 27, 28]

1 First try to engage the patient and establish an alliance so they are willing to try treatment. This can be difficult to accomplish, as many patients are delusional, prefer cosmetic treatment, are rejection sensitive, and do not want other people (including a clinician) to see them.
2 Empathize with the patient's suffering.
3 Take patients' appearance concerns seriously, neither dismissing their concerns about how they look nor agreeing that there is something wrong with their appearance. Trying to convince patients (especially delusional patients) that their beliefs are irrational or that they look normal is usually not helpful.
4 Instead, focus on the potential for psychiatric treatment to diminish their distress and preoccupation and improve their functioning and quality of life.
5 Provide psychoeducation about BDD and recommend reading.
6 For patients who wish to pursue cosmetic treatments, explain that such treatment appears ineffective for BDD.
7 Provide education about recommended treatments. It can be helpful to explain, for example, that SRIs are usually well tolerated, are not habit forming, appear to normalize the brain (and do not cause brain damage), and often diminish suicidal thinking in people with BDD. CBT is a practical 'here-and-now' treatment in which patients actively collaborate with the therapist and learn helpful skills by attending sessions and doing homework.

Treatment should be initiated with an SRI and/or CBT. SRIs are also the first-line medication for delusional BDD. All severely ill patients, especially those who are highly suicidal, should, in the author's opinion, receive an SRI. Patients with severe comorbid depression also warrant SRI treatment. Other comorbidity may warrant additional medication.

Before concluding that an SRI is ineffective, it should be tried for 12–16 weeks, reaching the highest dose recommended by the manufacturer or tolerated by the patient (if necessary) for at least 2–3 of those 12–16 weeks. If tolerated, higher doses than those recommended by the manufacturer can be cautiously tried to obtain or optimize a response (excluding clomipramine). If this is not effective, an augmentation strategy or switching to another SRI is indicated. For patients who refuse or are ambivalent about treatment motivational techniques should be tried. Patients who are not improving with CBT may need more frequent sessions, longer sessions, or a change in the current CBT focus. At least 4–6 months of weekly or more frequent CBT sessions, plus daily homework, is generally recommended. More severely ill and delusional patients may require much longer or more intensive treatment. Maintenance/booster sessions following CBT may reduce relapse risk. It may be helpful to add CBT to an SRI, or vice versa, if either treatment alone is insufficient.

For certain patients, adjunctive individual supportive therapy or family therapy may be helpful.[3] Families can be an invaluable support and facilitate treatment.[3] Mental health professionals may need to interface with dermatologists, plastic surgeons, and other physicians from whom patients have requested or are receiving cosmetic treatment.

Prevention

Because little is known about BDD's pathogenesis and risk factors, the disorder cannot currently be prevented. However, BDD should be treated in its early stages, before it causes substantial morbidity or interferes with a child's, adolescent's, or young adult's development.

Conclusions

Although this disorder has received far less investigation than many other serious mental illnesses, BDD research is rapidly advancing. It is important that clinicians screen patients for this often-secret disorder and be aware that it typically goes unrecognized in clinical settings, causing significant morbidity. While much more research is needed, available treatments are often very helpful for this distressing and often-disabling disorder.

Further information

Further information about BDD and its treatment is provided in references 3, 27, and 28 below and at www.bodyimageprogram.com.

References

1. Morselli, E. (1891). Sulla dysmorphophobia e sulla tafefobia. *Bolletinno della R accademia di Genova*, **6**, 110–9.
2. Phillips, K.A. (1991). Body dysmorphic disorder: the distress of imagined ugliness. *The American Journal of Psychiatry*, **148**, 1138–49.
3. Phillips, K.A. (1996, 2005). *The broken mirror: understanding and treating body dysmorphic disorder*. Oxford University Press, New York. Revised and Expanded Edition, 2005.
4. Phillips, K.A., McElroy, S.L., Keck, P.E. Jr., *et al.* (1993). Body dysmorphic disorder: 30 cases of imagined ugliness. *The American Journal of Psychiatry*, **150**, 302–8.
5. Phillips, K.A., Menard, W., Fay, C., *et al.* (2005). Demographic characteristics, phenomenology, comorbidity, and family history in 200 individuals with body dysmorphic disorder. *Psychosomatics*, **46**, 317–32.
6. Fontanelle, L.F., Telles, L.L., Nazar, B.P., *et al.* (2006). A sociodemographic, phenomenological, and long-term follow-up study of patients with body dysmorphic disorder in Brazil. *International Journal of Psychiatry in Medicine*, **36**, 243–59.
7. Phillips, K.A., Menard, W., Fay, C., *et al.* (2005). Psychosocial functioning and quality of life in body dysmorphic disorder. *Comprehensive Psychiatry*, **46**, 254–60.
8. Hollander, E., Cohen, L.J., and Simeon, D. (1993). Body dysmorphic disorder. *Psychiatric Annals*, **23**, 359–64.
9. Veale, D., Gourney, K., Dryden, W., *et al.* (1996). Body dysmorphic disorder: a survey of fifty cases. *The British Journal of Psychiatry*, **169**, 196–201.
10. Perugi, G., Akiskal, H.S., Giannotti, D., *et al.* (1997). Gender-related differences in body dysmorphic disorder (dysmorphophobia). *The Journal of Nervous and Mental Disease*, **185**, 578–82.
11. Phillips, K.A., Menard, W., Pagano, M., *et al.* (2006). Delusional versus nondelusional body dysmorphic disorder: clinical features and course of illness. *Journal of Psychiatric Research*, **40**, 95–104.
12. Phillips, K.A. (2004). Psychosis in body dysmorphic disorder. *Journal of Psychiatric Research*, **38**, 63–72.
13. Phillips, K.A. and Diaz, S. (1997). Gender differences in body dysmorphic disorder. *The Journal of Nervous and Mental Disease*, **185**, 570–7.
14. Phillips, K.A., Coles, M., Menard, W., *et al.* (2005). Suicidal ideation and suicide attempts in body dysmorphic disorder. *The Journal of Clinical Psychiatry*, **66**, 717–25.
15. Phillips, K.A. and Menard, W. (2006). Suicidality in body dysmorphic disorder: a prospective study. *The American Journal of Psychiatry*, **163**, 1280–2.
16. Cotterill, J.A. and Cunliffe, W.J. (1997). Suicide in dermatological patients. *The British Journal of Dermatology*, **137**, 246–50.
17. Buhlmann, U., Etcoff, N.L., and Wilhelm, S. (2006). Emotion recognition bias for contempt and anger in body dysmorphic disorder. *Journal of Psychiatric Research*, **40**, 105–11.
18. Buhlmann, U., Wilhelm, S., McNally, R.J., *et al.* (2002). Interpretive biases for ambiguous information in body dysmorphic disorder. *CNS Spectrums*, **7**, 435–6, 441–3.
19. Sarwer, D.B. (2002). Awareness and identification of body dysmorphic disorder by aesthetic surgeons: results of a survey of American Society for Aesthetic Plastic Surgery members. *Aesthetic Surgery Journal*, **22**, 531–5.
20. Phillips, K.A., Menard, W., and Fay, C. (2006). Gender similarities and differences in 200 individuals with body dysmorphic disorder. *Comprehensive Psychiatry*, **47**, 77–87.
21. Phillips, K.A. and Hollander, E. (in press). Treating body dysmorphic disorder with medication: Evidence, misconceptions, and a suggested approach. *Body Image: An International Journal of Research.*
22. Grant, J.E., Kim, S.W., and Crow, S.J. (2001). Prevalence and clinical features of body dysmorphic disorder in adolescent and adult psychiatric inpatients. *Journal of Clinical Psychiatry*, **62**, 517–22.
23. Rief, W., Buhlmann, U., Wilhelm, S., *et al.* (2006). The prevalence of body dysmorphic disorder: a population-based survey. *Psychological Medicine*, **36**, 877–85.
24. Deckersbach, T., Savage, C.R., Phillips, K.A., *et al.* (2000). Characteristics of memory dysfunction in body dysmorphic disorder. *Journal of the International Neuropsychological Society*, **6**, 673–81.
25. Saxena, S. and Feusner, J.D. (2006). Toward a neurobiology of body dysmorphic disorder. *Primary Psychiatry*, **13**, 41–8.
26. Phillips, K.A., Pagano, M.E., Menard, W., *et al.* (2006). A 12-month follow-up study of the course of body disorder. *The American Journal of Psychiatry*, **163**, 907–12.
27. National Collaborating Centre for Mental Health. (2006). Core interventions in the treatment of obsessive compulsive disorder and body dysmorphic disorder (a guideline from the National Institute

for Health and Clinical Excellence, National Health Service). British Psychiatric Society and Royal College of Psychiatrists, London. http://www.nice.org.uk/page.aspx?o=289817
28. Wilhelm, S., Phillips, K.A., and Steketee, G (in press). *Cognitive-behavioral therapy for body dysmorphic disorder: a modular treatment manual*. Guilford Publications Inc., New York.
29. Neziroglu, F. and Khemlani-Patel, S. (2002). A review of cognitive and behavioral treatment for body dysmorphic disorder. *CNS Spectrums*, **7**, 464–71.
30. Crerand, C.E., Franklin, M.E., and Sarwer, D.B. (2006). Body dysmorphic disorder and cosmetic surgery. *Plastic and Reconstructive Surgery*, **118**, 167e–80e.

5.2.9 Factitious disorder and malingering

Christopher Bass and David Gill

Factitious disorder

Introduction

Patients with factitious disorder feign or simulate illness, are considered not to be aware of the motives that drive them to carry out this behaviour, and keep their simulation or induction of illness secret. In official psychiatric nomenclature, factitious disorder has replaced the eponym Munchausen syndrome, introduced by Asher[1] to describe patients with chronic factitious behaviour. Asher borrowed the term from Raspe's 1785 fictional German cavalry officer, Baron Karl von Munchausen, who always lied, albeit harmlessly, about his extraordinary military exploits.

The criteria for factitious disorder in DSM-IV[2] are (a) the intentional production or feigning of physical or psychological signs or symptoms; (b) motivation to assume the sick role; and (c) lack of external incentives for the behaviour (e.g. economic gain, avoidance of legal responsibility, or improved physical well-being, as in malingering) and lack of a better classification for the disorders.

In the last 10 years there has been increased interest in deception in medical practice, with specific focus on pathological lying and the diagnostic dilemmas in this field: specifically, how to differentiate between hysteria, factitious disorders, and malingering. Some of these topics will be discussed in the next section.

This chapter concentrates on factitious physical complaints; fabricated psychological symptoms are considered under malingering.

Diagnostic problems

The DSM-IV criteria have recently come under attack. Turner[3] has argued that criterion B (motivation to assume the sick role) has no empirical content and fulfils no diagnostic function. He also argues that criterion A, the intentional production of physical or psychological signs or symptoms, emphasizes symptoms and cannot accommodate pseudologia fantastica (PF), voluntary false confessions, and impersonations. He concludes that the two criteria need reformulating in terms of lies and self-harm, respectively. Bass and Halligan[4] have also suggested that because the conceptual justification for factitious disorders is 'empirically unsubstantiated' and the motivation for diagnostic purposes (conscious versus unconscious; voluntary versus involuntary) essentially unknowable, it seems reasonable to question the clinical status and legitimacy of factitious disorder. More recently there has been a resurgence of interest in **pathological lying**, because this is often easier to identify than, for example, the degree of 'voluntariness' or 'motivation' to attain the sick role (however that is defined).

Pathological lying (pseudologia fantastica): a key component of factitious disorder

It is possible to identify pathological lying if the clinician has sufficient information at his disposal (most often the medical notes). If the patient reports, for example, that they are being treated for leukaemia, and when there is evidence that contradicts this, then this suggests dissimulation. On some occasions the patient will admit to lying, but this is rare.

Because pathological lying is often a key component in factitious disorders, evidence for it should be actively sought by the clinician. But what distinguishes the pathological liar from the person who just lies a lot? Dike *et al.*[5] suggest that the diagnosis is made when lying is persistent, pervasive, disproportionate, and not motivated primarily by reward or other external factors. They also suggest, however, that a key characteristic of pathological lying may be its compulsive nature, with pathological liars 'unable to control their lying'. Psychiatric conditions that have been traditionally associated with deception in one form or another include malingering, confabulation, Ganser's syndrome, factitious disorder, borderline personality disorder and antisocial personality disorder. Lying may also occur in histrionic and narcissistic personality disorders. It is important to note however that pathological lying can occur in the absence of a psychiatric disorder, and that there may be different types of pathological lying, e.g. the benefit fraudster and the stereotypical wandering Munchausen patient describe different subgroups. Furthermore, it has been reported that up to 40 per cent of cases of pseudologia fantastica have a history of central nervous system abnormalities, which suggests that brain dysfunction in these patients requires closer study.[6]

In recent years, functional neuroimaging techniques (especially functional magnetic resonance imaging) have been used to study deception. Attempted deception is associated with activation of executive brain regions (particularly prefrontal and anterior cingulate cortices), while truthful responding has not been shown to be associated with any areas of increased activation (relative to deception).[7] Furthermore, Yang *et al.*[8] reported that pathological liars showed a 22–26 per cent increase in prefrontal white matter and a 36–42 per cent reduction in prefrontal grey/white ratios compared with both antisocial controls and normal controls. These findings suggest that increased prefrontal white matter developmentally provides a person with the cognitive capacity to lie, although Spence[9] has urged caution in the interpretation of these results.

Clinical features of factitious disorder

Clinical features are diverse, and attempts to subtype patients have not always been helpful. The majority of patients with factitious disorders are non-wandering, socially conformist young women (often nurses) with relatively stable social networks.[10–12] These patients

are likely to enact their deceptions in general hospitals, especially accident and emergency departments, and the liaison psychiatrist should be alert to these clinical problems, which can be referred from a variety of different medical and surgical specialties.

Factitious disorders typically begin before the age of 30 years;[13,14] there are often prodromal behaviours in childhood and adolescence (see below). These individuals often report an unexpectedly large number of childhood illnesses and operations, and many have some association with the health care field.[10] High rates of substance abuse, mood disorder, and borderline personality disorder have been reported.[10,12] Approximately four-fifths of factitious disorder patients are women, and 20–70 per cent work in medically related occupations.[12]

Clinicians should be alert to the presentation of more exotic forms of factitious presentation. For example, some women present themselves to family cancer or genetic-counselling clinics and provide a false family history of breast cancer to their medical attendants.[15] Another recent example is 'electronic' factitious disorder,[16] used to describe patients who falsify their electronic medical records to create a factitious report (e.g. of cancer). Another important group is encountered in pregnancy, and this will clearly have important implications for child protection.[17]

In medico-legal practice factitious disorders have been described in patients with a diagnosis of reflex sympathetic dystrophy (RSD), specially involving the forearm,[18] and others have reported that the abnormal movements commonly associated with RSD (CPRS Type I) are consistently of somatoform or malingered origin.[19] Cases have been described where patients involved in litigation have died of factors directly related to factitious physical disorder.[20]

It is being increasingly recognized that these disorders can occur in childhood and adolescence, and child psychiatrists need to be alert to factitious presentations, especially in departments of infectious diseases.[21] Unlike adult patients, many of these children admit to their deceptions when confronted, and some have positive outcomes at follow-up. The descriptions of some of these children as bland, depressed, and fascinated with health care are remarkably similar however to adults with factitious disorders.[22]

Classification

Four main subtypes are distinguished in DSM-IV.[2]

1 Factitious disorder with predominantly psychological signs and symptoms. This is more difficult to diagnose than factitious disorder with physical complaints, because there is no way of excluding a 'true' psychiatric disorder by physical examination or laboratory investigation: see below under malingering.

2 Factitious disorders with predominantly physical signs and symptoms. Almost every illness has been produced factitiously. However, four subgroups describe most cases[23]:

   (a) self-induced infections

   (b) simulated illnesses, for example adding blood to urine

   (c) interference with pre-existing lesions or wounds

   (d) surreptitious self-medication, for example self-injection of insulin

   These categories are not mutually exclusive or jointly exhaustive.

3 Factitious disorders with combined psychological and physical symptoms

4 Factitious disorders not otherwise specified. This includes factitious disorder by proxy (see below and Chapter 9.3.3).

Diagnosis

Clinicians should become suspicious that a patient may be fabricating symptoms if the following features are noted:

- The course of the illness is atypical and does not follow the natural history of the presumed disease, e.g. a wound infection does not respond to appropriate antibiotics (self-induced skin lesions often fall into this category, when 'atypical' organisms in the wound may alert the physician).
- Physical evidence of a factitious cause may be discovered during the course of treatment, e.g. a concealed catheter, a ligature applied to a limb to induce oedema.
- The patient may eagerly agree to or request invasive medical procedures or surgery.
- There is a history of numerous previous admissions with poor outcome or failure to respond to surgery (these patients may overlap with the chronic somatoform patient with 'surgery prone behaviour').[24]
- Many physicians have been consulted and have been unable to find a relevant cause for the symptoms.

Additional clues include the patient being socially isolated on the ward and having few visitors, or the patient being prescribed (or obtaining) opiate medication, often pethidine, when this drug is not indicated. When these findings occur in someone who has either worked in or is related to someone who has worked in the health field, the caregivers should have a high index of suspicion for a factitious disorder. Obtaining collateral information from family members, prior physicians, and hospitals is crucial.

Differential diagnosis

Factitious disorder must be distinguished from authentic medical conditions. It is not uncommon in clinical practice however to find patients with **both** factitious disorder and coexisting physical illness. For example, patients with brittle diabetes are usually young females who deliberately interfere with their treatment, causing unstable diabetic control.[25] A syndrome of severely unstable asthma ('brittle asthma') which also affects young females has also been described;[26] this can occur (especially in A and E departments) with paradoxical adduction of the vocal cords during inspiration.[27] Such patients can neglect to take medication at appropriate times and then ignore adequate management of the potentially dangerous consequences. This may lead to repeated admissions to hospital with medical emergencies such as diabetic ketoacidosis, status asthmaticus, or even pseudo status (simulated status epilepticus).

Factitious disorders are differentiated from somatoform disorders, where physical symptoms, although not caused by physical disease, are deemed not to be intentionally produced. Patients with factitious disorder, although they may state that they are not aware of the motives that drive them, voluntarily produce their physical or psychological symptoms. The disorders may overlap, however, and non-wandering female patients with factitious physical disorders may have more in common with those females who have somatoform disorders than with men with factitious disorder or with malingerers. Fink[28] found that 20 per cent of patients with

persistent somatization (i.e. patients with more than six admissions to the general hospital with medically unexplained symptoms) also had a factitious illness. One of the authors (Christopher Bass) has also found coexisting chronic somatoform and factitious disorders in the female perpetrators of factitious or induced illness.[29]

A more difficult distinction is between factitious disorder and malingering. Malingerers, described below, have clear-cut goals, often personal profit, and lack a history of hazardous, unnecessary invasive procedures. In our opinion the boundaries between the two disorders are more porous than the glossaries would have us believe, and we have stated above that the differentiation between conversion disorders, factitious disorders and malingering is extremely difficult in clinical practice. Case reviews have demonstrated how behaviour may shift from somatization to factitious to malingering when patients are followed longitudinally.[30] In our opinion the clinical status and diagnostic legitimacy of factitious disorder as a selective medical disorder is questionable, because it fails to take account of a morally questionable but volitional-based choice to deceive others by feigning illness. Considered as an act of wilful deception, illness deception can be meaningfully conceptualized within a socio-legal or moral model of human nature that recognizes the capacity for choice and the potential for pursuing benefits associated with the sick role. This model, which recognizes the human capacity to exercise free will, is shown diagrammatically in Fig. 5.2.9.1.

Epidemiology

Factitious disorders are relatively uncommon but probably underdiagnosed. Prevalence depends on clinical setting and the investigators' index of suspicion. Factitious disorder (probably underreported) is probably more common than full-blown Munchausen syndrome (probably overreported). In a recent survey of physicians from Germany, frequency estimates of factious disorder among their patients averaged 1.3 per cent, with dermatologists and neurologists giving the highest estimations.[31]

Of 1288 liaison psychiatric consecutive referrals, seen in a North American general hospital, 0.8 per cent had factitious disorder.[13] Similar figures have been reported by Dutch investigators.[14]

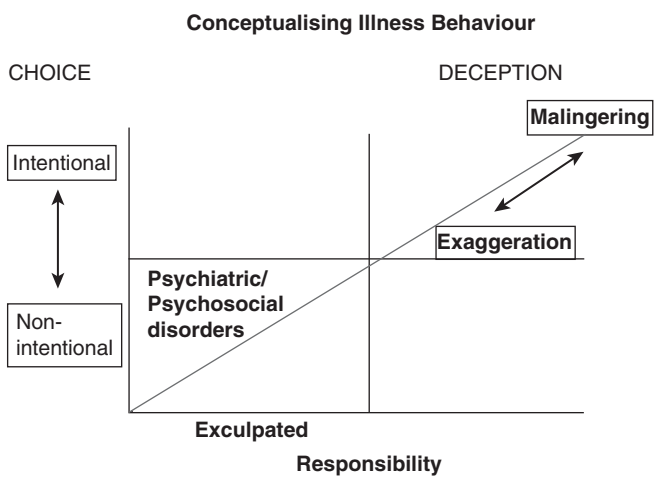


Fig. 5.2.9.1 Model of illness deception incorporating patient choice, free will and intentionality. (Reproduced from Bass, C. and Halligan, P. Illness related deception: social or psychiatric problem? Journal of the Royal Society of Medicine, **100**, 81–4. Copyright 2007, The Royal Society of Medicine Press.)

Prevalence rates for factitious disorder with psychological symptoms in patients under the age of 65 in a psychiatric hospital are approximately 0.4 per cent.[32]

Aetiology

There is little aetiological information, as large studies are lacking and the self-reported histories of many patients are fallacious. However, a number of themes are apparent[13]:

Developmental

- Parental abuse, neglect, or abandonment: Many factitious disorder patients experienced significant deprivation in childhood that left them with unfulfilled cravings for attention and care. These patients then seek to gratify these dependency needs by creating illness to obtain the 'attention and care' of the medical system
- early experiences of chronic illness or hospitalization

Medically related

- a significant relationship with a physician in the past
- experiences of medical mismanagement leading to a grudge against doctors
- paramedical employment

Physical

- Organic brain disorder: There is increasing evidence that neurobiological factors have a role in some patients, and it has been recommended that screening for evidence of brain dysfunction be carried out in these patients.[33] It is thought however that it is the pseudologia fantastica, and not the factitious disorder per se, that is associated with brain dysfunction (see previous section on Pathological Lying), though such a distinction may be difficult to support in clinical practice.

The authors have found it useful to conceptualize the problem using a cognitive behavioural formulation suggested by Fehnel and Brewer.[34] This allows the assessors to examine for relevant developmental factors and recent life events (especially losses or threatened losses and separations; see Fig. 5.2.9.2).

Course and prognosis

Factitious disorder may be limited to one or more brief episodes, but is usually chronic. For example, of 10 patients identified in a general hospital setting and followed up, at least one was known to have died as a result of factitious behaviour 4 months after the index admission.[13] Only one of the remaining nine patients accepted psychiatric treatment after discharge from hospital. Other authors have, however, reported a less gloomy outcome.[10] Outcome may be determined by how patients are managed, once their deceptions become manifest. Regrettably, the unmasking of the disorder is often the end of the story. Psychological support following hospital discharge may be associated with improved outcome.[10] Non-wanderers with more stable social networks may have a better prognosis than wanderers.[14] Engaging a patient with factitious disorder in long-term psychological treatment occurs so rarely that it often becomes the subject of a case report. One such report, of a 20-year follow-up of a patient with factious disorder with a favourable outcome is, in the opinion of the authors, the exception to the rule.[35]

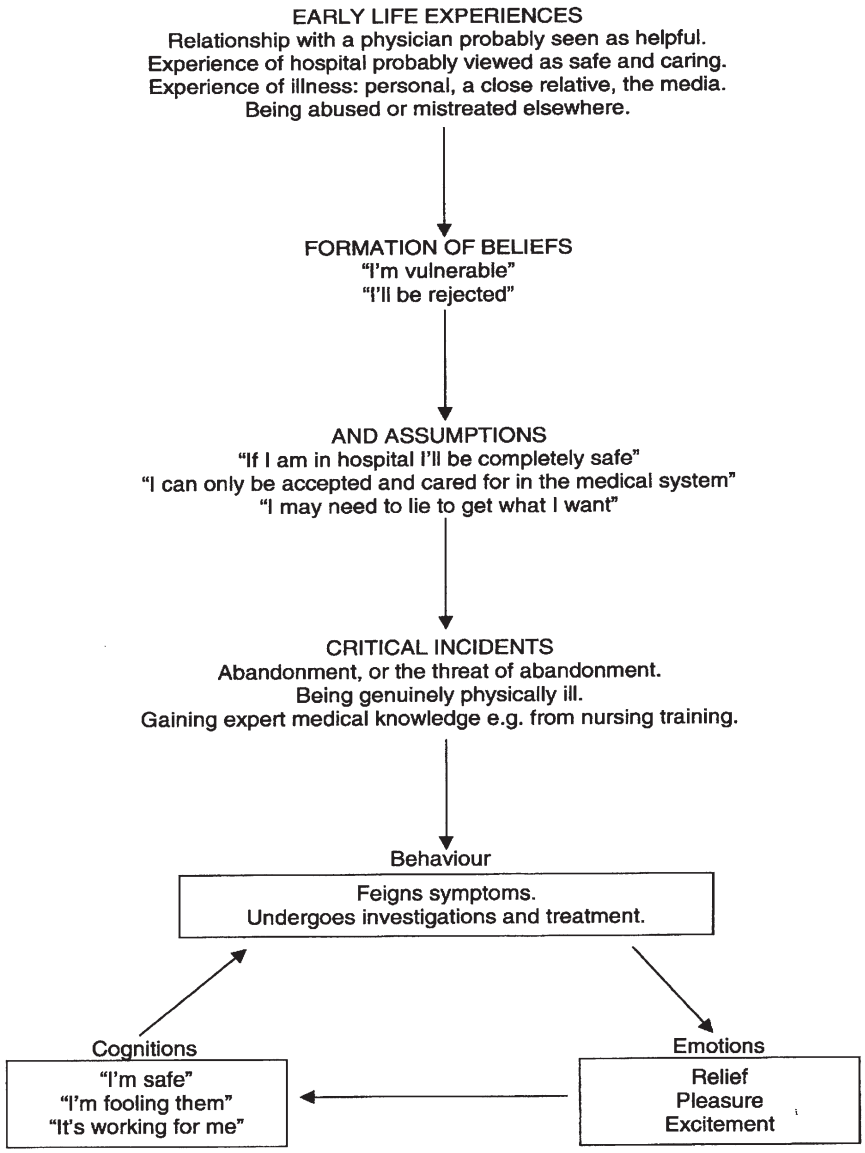


Fig. 5.2.9.2 A cognitive behavioural conceptualization of factitious disorder. Reproduced from Kinsella, P. Factitious disorder: a cognitive behavioural perspective, *Behavioural Cognitive Psychotherapy*; **29**: 195–202, copyright 2001, with permission from Cambridge University Press

Treatment

There are no systematic or controlled treatment studies on patients with factitious disorders. This is hardly surprising, as the patient's primary motive is deception, and the doctor's is to understand or unmask these motives, usually leading to rapid discharge from hospital after the deception has been exposed.

Management

Once the diagnosis is established, the doctor–patient relationship may have become irreparably damaged: negative emotions in the doctor may need to be dealt with before any consideration can be given to 'engaging' the patient in any therapeutic endeavour. Ethical and legal issues may also intrude (see below) and affect management. Although psychotropic medications have been used, the main treatment is psychological, using either confrontational or non-confrontational strategies.

Before treatment takes place however, it is important to establish the diagnosis, which is nearly always initially made by a non-psychiatrist, e.g. A and E physician, infectious disease specialist. A meeting should take place between the physician/surgeon and the psychiatrist and a strategy worked out before any confrontation or other approach is embarked on. These preliminary procedures are important and preparation before the joint interview is crucial (Table 5.2.9.1).

(a) Confrontational approaches

This process is easier if the physician has tangible evidence of fabrication, for example catheters, or medication used in the patient's deception. It is also desirable to have the psychiatrist present when

Table 5.2.9.1 Supportive confrontation: preparation and process (for non-psychiatrists)

- Collect firm evidence of fabrication, e.g. catheter, syringe, medication
- Discuss with psychiatrist (or hospital legal team if no psychiatrist available)
- Arrange meeting to marshall the facts; discuss strategy. Discuss with GP
- CONFRONTATION with patient should be non-judgemental, non-punitive, with . . .
- Proposal of ongoing support/follow-up
- If patient is a health care worker, the doctor should discuss with his/her medical defence organization
- Discuss the outcome of the confrontation with the patient's GP
- Document full record of the meeting and proposed outcome in patient's notes

the physician confronts the patient. The approach during confrontation and thereafter should be non-punitive and supportive, stressing continuity of care, and that the patient is a sick person who needs help. This approach was adopted in perhaps the largest published series of patients with factitious disorder treated systematically.[10] Thirty-three patients were 'confronted' with objects found in their room or with clinical data showing that their conditions were factitious. Only 12 (36 per cent) patients acknowledged the truth; the remaining 21 continued to deny that they played any role in creating their disorders. No confronted patient developed serious psychological disturbance or became suicidal, or discharged themselves against medical advice. Four of the most chronic cases became asymptomatic. Most, however, greeted the idea with either overt hostility or passivity and covert negativism.

More recently Krahn *et al.*[12] replicated these findings, but found that only one in six of their patients acknowledged their factious behaviour. Many patients will experience confrontation as humiliating and seek care from a different hospital. Others will refuse to see the psychiatrist with the treating clinician to discuss the deception, and discharge themselves against clinical advice.

(b) Non-confrontational strategies

These approaches, advocated by Eisendrath and Feder,[36] are less concerned with the origin of the illness and more with shaping future behaviours. Face-saving is a key element, and it is important for the patient to subsequently explain their 'recoveries' without admitting that their original problems were psychiatric.

One strategy is the therapeutic 'double-bind'. In this approach the patient is presented with two choices: prove that his or her disorder is not factitious by responding to a relatively minor and benign medical intervention, or prove that the disorder is factitious by failing to respond. For example, a woman was offered the double-bind for a wound that had failed to heal in 4 years despite numerous surgical closures. Following this strategy the plastic surgeon told her that her wound should respond to a skin grafting procedure. If it did not, it would mean that her disorder was factitious in origin. The graft took place, and there was no recurrence of infection at 2-year follow-up.[36] This approach has also been used with some success in the rehabilitation of three patients with factitious motor disorders.[37] The strategy was successful in providing patients with a face-saving legitimization of both their illnesses and recoveries.

Another face-saving approach uses 'inexact interpretations', i.e. suggesting a relationship between certain events or stressors, for example being abandoned, and emergence of factitious symptoms. It involves presenting a brief formulation of the problem to the patient, stopping short of overtly identifying the factitious origin. By avoiding confrontation the doctor makes it safe for the patient to relinquish the symptom with a feeling of control. Regrettably, none of these non-confrontational techniques have been evaluated in a systematic fashion.

(c) Systemic interventions

Patients with factitious disorders can create havoc on medical and surgical wards. They often elicit negative and hostile emotions in general hospital staff, especially after the deception has been exposed. The psychiatrist can help staff members to vent and reduce the anger they experience when a factitious diagnosis is confirmed, and also help the staff to understand the likely mechanisms underlying the factitious behaviour. These issues are often best addressed at a multi-disciplinary staff meeting. The major task of this group, which should include a member of the hospital medico-legal department as well as the patient's family doctor, is to develop practical treatment guidelines and to discuss the complex legal and ethical issues raised with factitious physical disorders. Some of these issues are discussed in the next section.

Ethical and legal issues

Patients with factitious disorders create unique ethical and medico-legal issues, some of which will be described below.

(a) Confidentiality

If no meaningful doctor–patient relationship exists or can be established, it has been argued that the physician is not bound by ethical codes, and that drastic solutions such as keeping 'blacklists' and the use of a central register can therefore be justified.[38] Objections to these approaches include breach of doctor–patient confidentiality and possible denial of treatment for genuine illness. Anyway, the use of aliases and poor record-keeping reduces the effectiveness of blacklists. Furthermore, physicians who disclose information without patient consent may have to justify the decision to their licensing body. In the United States, there is a consensus that disclosure should only occur where there is a specific risk to the patient and/or another party. In such situations, a multi-disciplinary staff meeting can help to develop treatment policy, and share responsibility for difficult decisions (see above).

(b) Invasion of privacy

The medical literature contains many descriptions of how the diagnosis of factitious disorder was established following a search of the patient's room or belongings. Some physicians, however, consider that such behaviour infringes patients' rights, and that no search should be undertaken without the patient's knowledge and consent. One way of avoiding this dilemma is to make it clear to the patient that factitious disorder is among the differential diagnoses, and then request permission for a room search. If needles or syringes are discovered during the course of treating the patient, the ethical issue of invasion of privacy does not arise.

(c) Involuntary hospitalization or treatment

Because the patient with factitious disorder may engage in behaviour that leads to permanent maiming or even death, it has been argued that in such cases a compulsory order may be used to protect the patient from himself or herself. This will provide time for

not only a more in-depth psychiatric assessment but also the development of a more trusting relationship with a therapist. This is a contentious subject, but some case reports do indicate that extended involuntary hospitalization may result in therapeutic progress.[39]

Induced factitious illness (Munchausen syndrome by proxy)

Meadow[40] first described this disorder as 'the deliberate production or feigning of physical or psychological symptoms or signs in another person who is under that individual's care', which has recently been renamed fabricated or induced illness or FII.[29] The perpetrator is usually the mother and the victim her child: the syndrome is considered to be a form of child abuse. The parent's aim is to have the child considered seriously ill: this may involve providing false histories, poisoning, or persuading doctors to carry out invasive and potentially dangerous procedures.

Factitious disorder and FII can be interrelated. Psychiatrists (and general practitioners) should be aware of the implications of a diagnosis of factitious disorder for any children of the index patient. For example, 75 per cent of mothers of these children have a history of a factitious or somatoform disorder, and most meet criteria for personality disorder.[41] With the birth of a child some mothers with pre-existing factitious disorder abandon dissimulation themselves, only to extend it to the next generation through factitious disorder by proxy. Because factitious disorder and FII can co-occur, the finding of one should always trigger judicious efforts to establish, or hopefully disconfirm, the other.

Long-term outcome in factitious disorder by proxy is poor, so active early intervention is recommended, with child protection agencies working closely with paediatric and psychiatric services (see also Chapter 9.3.3).

Factitious disorder in health care workers

If factitious disorder is diagnosed in a health care worker, the investigating psychiatrist should consider whether their continuing clinical work would pose risks to either patients (often children) with whom that person comes into contact (i.e. of factitious disorder by proxy), and/or the health of the factitious disorder patient himself or herself. These issues have been thrown into sharp focus by some highly publicized cases in both the United States and the United Kingdom.[42,43] Any employee who manufactures crises, for example multiple cardiac arrests and resuscitations, is obviously of great concern. The British commission that investigated the 1993 case made several stringent recommendations, for example 'We recommend that no candidate for nursing in whom there is evidence of major personality disorder should be employed in the profession'.[44] Nevertheless, psychiatrists should seek medico-legal advice before communicating concerns about such patients to any third party, including the employing hospital.

Malingering and exaggeration

Introduction

Malingering is the deliberate simulation or exaggeration of physical or psychiatric symptoms for obvious and understandable gain (e.g. financial compensation, disabled status and benefits, avoidance of criminal prosecution or conscription).

Treating doctors have been understandably reluctant to diagnose malingering lest it adversely affect the patient. However, recent research indicates that it may not be uncommon, especially where financial rewards attach to disability status, such as benefits for sickness or compensation for injury. We here describe some types of malingering seen in psychiatry, and discuss newer specialist psychological tools, especially *symptom validity testing*, which are emerging as useful additions to the overall assessment.

Definition

Malingering is not coded as a mental disorder in either ICD-10 or DSM-IV, although in the latter it is denoted as an 'additional condition that may be the focus of clinical attention'.

DSM-IV[2] suggests that malingering should be 'strongly suspected' when two or more of the following factors apply:

1 medico-legal context

2 antisocial personality disorder

3 discrepancy between complaints and objective findings

4 lack of co-operation with the assessment

This actually sets a very low bar: in effect, DSM advises it should be 'strongly suspected' in any disputed medico-legal case (where points 1 and 3 apply), but does not give guidance as to how this 'suspicion' should be followed up. The ICD-10 definition ('the intentional production or feigning of either physical or psychological symptoms or disabilities, motivated by external stresses or incentives') is broad, but again does not give practical suggestions regarding assessment.

Malingering includes:

1 Pure malingering: complete fabrication of symptoms

2 Partial malingering: exaggerating real symptoms or saying that past symptoms are continuing

3 False attribution: falsely saying that real health problems are due to a compensable accident or other circumstance

Epidemiology

Until recently, there has been little systematic information on prevalence, although it has generally been agreed that exaggeration of real symptoms is more common than outright faking expect them to dissimulate. Exaggeration, dissimulation, or feigning can be considered one of several rational/economic/adaptive options open to patients when seeking health care and/or limited social and welfare resources.

Using the 'Composite Disability Malingering Index' Griffin *et al.*[45] suggested that 19 per cent of disability claimants in the United States malingered to some degree. A recent study of 131 practicing members of the American Board of Clinical Neuropsychology provided estimates of the prevalence of malingering and symptom exaggeration for a variety of different clinical conditions.[46] In this study, estimates of the base rate of malingering/symptom exaggeration were calculated using over 33 000 annual cases seen by a group of clinical neuropsychologists. The reported base rates (when statistically adjusted to remove for the influence of referral source) were 29 per cent for personal injury, 30 per cent in the case of disability or workers compensation, 19 per cent in criminal cases, and 8 per cent in medical or psychiatric cases. The same rates

broken down by diagnosis revealed 39 per cent in the case of mild head injury, 35 per cent in fibromyalgia and chronic fatigue, 31 per cent in chronic pain, 15 per cent for depressive disorders, and 11 per cent in the case of dissociative disorders. In a separate review of 1363 compensation-seeking cases, Larrabee[47] found similar figures for mild head injury of 40 per cent. The use of symptom validity testing has confirmed these high rates of exaggeration (see under 'Psychological tests'). For example, about 42 per cent in the Canadian series of Richman *et al.*[48] and about 60 per cent in the United Kingdom in a series studied by one of the authors (Gill *et al.* submitted for publication).

Clinical features

(a) Malingered neurosis: post-traumatic stress disorder

The archetypal disorder after trauma is post-traumatic stress disorder. Malingering of other disorders seen after trauma, notably depression is less frequent, confirming malingerers' preference for dramatic positive symptoms such as nightmares and flashbacks.

In 1983 Sparr and Pankratz[49] described five men who claimed to have been 'traumatized' in the Vietnam War; three claimed to have been PoWs. It turned out that none had been PoWs, four had never been in Vietnam and two had never been in the services. The patients were seeking the generous benefits, which the United States accords to ex-service (Veterans' Administration) personnel. The key point in this paper was that the authors *supplemented their clinical assessment by seeking external data*, in this case service records; relying on clinical assessment alone would have led to wrong diagnosis. Seeking corroboration is vital in the assessment of possible malingering.

Rosen, in another classic paper, documented malingered PTSD in the case of the *Aleutian Enterprise*, a fish-processing ship, which sank in the Bering Sea in 1990.[50] Of the 31 on board, 9 were lost, 2 went back to sea, and the remaining 20 sued. Nineteen (86 per cent) of these 22 survivors consulted psychiatrists or psychologists with the key features of PTSD. But this is much higher than the expected rates: most individuals exposed to a traumatic event do not develop PTSD. Even if they do, it resolves over a few weeks or months in many cases, whereas here the claimants' symptoms did not show any tendency to resolve. Furthermore, they all had almost all of the classic features of the condition, in other words, they did not display the case to case variability, which would be expected in real individuals. Rosen documented that the patients had in fact 'shared symptoms' and had been 'coached' by their attorneys, some of whom had advanced the claimants money so they would not have to settle.

Most cases of suspected malingered post-traumatic stress disorder involve less dramatic civilian trauma, most commonly road accidents. Again, exaggeration of genuine symptoms is much more common than outright fabrication. For example, a person involved in a car accident, with apparently genuine phobic travel anxiety, may report that they have nightmares and 'flashbacks', but their description of these experiences lacks vividness on close enquiry. Holistic assessment is vital; if a minor accident is to be accepted as causing a severe psychiatric condition such as PTSD, there must be some evidence of pre-existing vulnerability. Otherwise, the apparent result will be disproportionate to the cause, and the possible influence of external incentives on symptom presentation will need to be considered.

(b) Malingered psychosis

This disorder can occur in various circumstances, for example in homeless persons wishing to obtain shelter in hospital, in previously psychotic inpatients whose discharge is imminent, in illegal migrants seeking to avoid deportation or in criminal defendants trying to avoid standing trial or to influence sentencing.

A (perhaps somewhat academic) distinction can be made between malingered and factitious psychosis, in that malingerers are conscious of their motivation, and their goal is not merely confined to gaining patient status. However, the following description covers both. It is 'positive' symptoms, which are usually mimicked (e.g. hallucinations). They are often dramatic or bizarre. Patients are keen to describe them at interviews, unlike, for example, most patients with schizophrenia. Symptoms are obvious during assessments, less so when the patient is unobserved. More subtle features of genuine psychosis, such as thought disorder and negative symptoms, are absent. Florid hallucinations may be unaccompanied by delusions, which would be unusual in genuine psychosis.

For example, Jaffe and Sharma[51] described nine defendants on serious criminal charges that developed uncommon symptoms such as coprophagia and 'seeing little green men'. Eight were judged to be malingering, and fit to plead, based on evidence such as the association between visual hallucinations and organic brain syndromes, of which there was no evidence on investigation. Malingering in forensic settings also includes feigned memory deficits, when isolated amnesia for an alleged crime would place malingering in the differential diagnosis. Forensic psychiatry is dealt with elsewhere in this book.

Ganser syndrome does not appear in current classifications. It was originally described in prisoners, as comprising confusion and so-called 'approximate answers' (or Vorbereiden: *Question*: for example: how many legs has a horse? *Answer*: three). If true confusion (diminished level of consciousness) is present, the diagnosis is the cause of this. Approximate answers may be seen in several conditions, including mental retardation, organic brain disorders, and malingering; again, the diagnosis will be the underlying condition. The term Ganser syndrome has now largely and rightly been dropped.

(c) Malingered cognitive deficit

Study of cognitive deficits following brain injury has recently led to advances in understanding of malingering, through the development of special neuropsychological tests, especially *effort testing*, to gauge the effort the patient brings to cognitive testing. These tests seem likely to have broader application than just brain injury assessment.

Discussion of the question of malingering post head injury has until recent years been rich in opinion, though comparatively light on facts. Miller's view,[52] long influential, was that many patients malingered their memory and other cognitive symptoms and that symptoms were in inverse proportion to injury severity and were only resolved with receipt of compensation. He used the term 'accident neurosis'. Mendelson, by contrast,[53] found that disability continued after settlement in many patients, and inferred from this that disability was generally not malingered.

The question is obviously not capable of being resolved scientifically without data. But such data is now becoming available. Recent findings however have supported Miller's original observations that embellishment rises as injury severity decreases in a

compensable context.[54] Moving forward from mere debate, the American Academy of Neuropsychologists recently published a consensus statement which concluded that 'Symptom exaggeration or fabrication occurs in a sizeable minority of neuropsychological examinees, with greater prevalence in forensic contexts', and that the use of effort testing is mandatory in neuropsychological assessments.[55]

In clinical assessment, immediate recall (e.g. digit span) is important, as even organic amnesic patients (e.g. Korsakoff's syndrome) perform normally; poor performance suggests that poor motivation or malingering should be among the differential diagnoses. Claims of being unable to remember personal information (e.g. name and birthday, also preserved in organic amnesia), yet having been able to come to the assessment independently, are highly suggestive of malingering, but seen only in gross cases. However, these 'bedside' tests are only a guide: psychiatric assessment of brain injury patients is not complete without quantitative assessment of cognitive function, including effort testing (see below).

(d) Malingered physical disease

This usually presents either as a referral to a liaison psychiatrist, or in a medico-legal context. A frequent example is a patient with post-injury back or neck pain who is involved in litigation or seeking disability payments. Often, some form of accident has undoubtedly occurred, so the potential for initial physical injury is not in doubt; but the length and severity of symptoms, disability, and distress may seem out of proportion, and raise the possibility of malingering.

In a seminal paper, Richman[48] administered effort testing to 106 people claiming injury or sickness benefits. Forty-five (42 per cent) failed. On one easy subtest, those who failed the effort test overall had a similar score on average to patients with dementia tested previously, even though none of them had a clinical diagnosis of dementia. Schmand et al.[56] found that 61 per cent of litigants after whiplash neck injury had evidence of underperformance on memory testing compared with 29 per cent of outpatient controls. The underperforming litigants scored as low as controls with definite evidence of closed head injury.

Classification

Malingering should be distinguished from factitious disorder, and from other syndromes such as hypochondriasis, other somatoform disorders, and conversion/dissociation disorder. These distinctions may involve difficult judgements such as how 'intentional' is the production of a symptom, or how 'genuine' it is. As an alternative, it has been suggested that such patients lie on a continuum between those in whom the production of symptoms is assumed to be wholly unconscious (conversion/dissociation disorder) and those in whom it is wholly conscious (malingering, factitious disorder).

However, use of the concept of the unconscious has to be very cautious when there are external incentives. It stretches credulity to think that a claimant would be conscious of there having been an accident so as to pursue litigation, and conscious that the outcome could include financial compensation, but somehow unconscious that presentation of symptoms could form a desired link between the two.

It is possible that the emergence of effort testing may cast new light on the area of unexplained physical symptoms. For example, the concept of somatoform disorders assumes that the symptoms are not consciously produced. However, if large-scale studies reveal that a substantial proportion of somatoform patients turn out to fail effort tests, or in other words to display evidence of conscious symptom exaggeration, then the concept of somatoform disorders may need to be re-examined.

Diagnosis and differential diagnosis

Doctors' training and culture rightly encourage the treating physician towards a generally trusting relationship with his patients. However, assumptions that patients' accounts are generally trustworthy and that malingering is rare are not appropriate if the main responsibility of the doctor lies elsewhere, for example, to the Court, if he is preparing an expert report.

Identifying ungenuine cases requires an enquiring approach, and the methodical use of all sources of information:

- awareness of the possibility of exaggeration or faking of symptoms
- neutral attitude
- open questions initially; use closed questions with caution
- unlikely questions (see below)
- mental state—changes appropriately as sensitive topics discussed?
- informants (but they may also have vested interests)
- observation—overt (e.g. in a ward) or covert (e.g. video surveillance)
- medical records (and legal documents if applicable)
- look for consistency of accounts
- standard psychometric tests—consistency of results across measures
- specialist instruments: symptom validity tests

The medical notes should be read, ideally before interview, especially the general practice records, and any discrepancies noted for specific enquiry. Legal assessments, case papers and previous reports should be studied.

In the clinical interview, a neutral attitude is essential; a confrontational approach, even if malingering is strongly suspected, may cause further exaggeration of symptoms. Open questions should be used at first. Closed questions should be avoided (e.g. 'Do you ever get nightmares where you seem to re-experience the accident?'). The careful use of unlikely questions can be useful. For example, in suspected malingered post-traumatic stress disorder the answer 'Yes' to the question 'Have you had any problems with colour vision since the accident?' would be suggestive, but would still need clarification with open questions. 'I am now colour-blind' might suggest malingering, but 'Red makes me nervous—it was a red car which crashed into me' might not.

Questions must be appropriate to the case, ideally prepared beforehand and introduced tactfully to prevent the patient feeling that the interviewer is attempting to 'catch him out'. The interviewer should look for the clinical characteristics of the particular malingered disorder in question.

Surveillance by video or other means may be used by lawyers or insurers, although is seldom initiated by clinicians, for ethical reasons (except perhaps in suspected factitious disorder by proxy—see Chapter 9.3.3).

Psychological tests

Since the last edition of this book, consensus has developed that symptom validity testing is essential in neuropsychology, that is, in the field primarily concerned with memory problems and brain injury. The tests in practical use are in fact specialized memory tests, and it is likely that their use will be extended from head injury cases to other cases in which memory complaints occur (such complaints are in fact very common in patients with pain or distress for any reason). We are about to discuss what currently appear to be the leading symptom validity tests in UK neuropsychology, the Word Memory Test[57] and the ToMM,[58] but first we mention some other tests by way of historical background.

Inconsistent patterns of response on standard instruments (e.g. high scores on one measure of depression and low on another) might be suggestive. However, this should not be overinterpreted, because there is substantial test–retest variation on many tests, and there is also variation between subscales on the same occasion. Certain subscales of the Minnesota Multiphasic Personality Index have been proposed as measures of tendency to malinger. Specialist instruments have been developed, for example the Structured Inventory of Malingered Symptoms,[59] which identifies features associated with malingering such as endorsement of rare symptoms, which occur only infrequently in clinical populations. However, consensus on such tests remains some way off and they are not in general use in the United Kingdom.

Tests for malingered memory deficits present memory tasks, which appear difficult but are in fact easy. For example, 50/50 psychiatric and 10/16 mentally retarded inpatients were able to recall nine of the 15 items on the Rey 15-item.[60] Forced choice testing (e.g. Portland Digit Recognition Test) is another approach. A sequence of digits is presented. Subjects must identify it among two further sequences, one identical and the other different. By chance alone, they must score around 50 per cent on a large number of items, and so below chance scores (below 50 per cent) strongly suggest malingering.[61] However, this is very gross, and will miss many cases where the degree of exaggeration is less. Again, these tests are not in general use in the United Kingdom.

In the United Kingdom, the Word Memory Test and the ToMM have become established as leading symptom validity tests in neuropsychology. They are regarded as having the best research support and are administered in standard formats.

TOMM stands for *Test Of Malingered Memory*.[58] It is a pictorial test, not computerized, and is mainly used by psychologists. Word Memory Test (WMT) (Green: wordmemorytest.com) is available either on paper or on computer. It is mainly used by psychologists as part of a 'battery' of neuropsychological tests. An abbreviated form for physicians, the Medical Symptom Validity Test, is available.[57] The key point is that patients are given a test of memory, which looks difficult, but is in fact known to be easy from previous administration to control subjects.

Someone making a **good effort:**

- scores well on tests which are in fact easy (even though they may look hard)
- scores lower on more difficult tests

Someone making an **inconsistent or poor effort**:

- may score low on tests which look hard (though they are in fact easy)
- may not score lower on more difficult tests

If the patient's scores are low, where even say primary school children score almost perfectly, it would suggest that he might not have been making a full effort on the test. If at the end of the test he said that he had made a full effort, this would be evidence that his self-report of effort brought to cognitive testing is not accurate. This would be consistent with the proposition that he might have been exaggerating at least the memory aspects of his complaints, and by extension, possibly other symptom areas also.

Aetiology

Malingering is not a mental disorder, so complex general theories about causation are unlikely to be helpful. It has been suggested for example that[62] 'adaptation' is the simplest model for malingering: ' . . . malingering is more likely to occur when the evaluation is perceived as adversarial, when the personal stakes are very high, and when no alternatives appear to be viable'. However, this really just restates the problem in different terms. Nor do the 'response styles' described by psychologists seem to offer fundamental insights. It is more helpful to consider each case in a common sense way, bearing in mind the presence or absence of external incentives, and the results of the holistic assessment process outlined above, in combination with the results of symptom validity testing.

Course and prognosis

Malingerers are so heterogeneous that it is impossible to state prognosis in general. There are few adequate follow-up studies.

In forensic and inpatient settings, malingering is usually episodic, associated with particular circumstances such as impending discharge, trial, or change in conditions of imprisonment (e.g. transfer to a single cell or hospital wing). The behaviour often stops when the circumstances no longer remain, although they may recur. Similar situations occur in persons facing conscription, or in migrants at risk of repatriation.

The prognosis of malingering *in personal injury litigants* is unknown. Studies indicating poor outcome of many compensated litigants cannot be extrapolated, as the number of malingerers in these samples is unknown. Clinical experience suggests that patients with long-standing disability, even if partly or wholly non-organic, often fail to recover fully in any event.

Treatment

There is very little evidence on management, which will largely be dictated by whether the clinician has clinical responsibility, or whether he or she has been asked to give an opinion to a third party. Even if the diagnosis of malingering is clear, it may be appropriate to inform the referrer, rather than the patient directly, as the patient may become very angry. If there is a psychiatric disorder present when, as it were, the dust has settled, then this should be treated in standard fashion.

Clinical experience is that patients for whom the exercise of attempted malingering seems worthwhile do often have substantial pre-existing psychosocial problems, including lack of skills or employment; there is a tendency towards a regionality of such claims, with areas of deprivation ('rust-belt') overrepresented.

Efforts at retraining and vocational rehabilitation may be more likely to be of assistance in the long-term than specialist psychiatric care.

Possibilities for prevention

These seem mainly confined to malingering after injury. First, systems of litigation should be expedited. Prolonged cases certainly make for exaggeration of symptoms and disability. Second, some patients with chronic disability, irrespective of cause, do respond to rehabilitative treatment (e.g. programmes assisting in return to work or in the management of chronic pain), even at a late stage. If such programmes were easily and generally available at an earlier stage in the evolution of the disorder, symptoms and subsequent disability could certainly be ameliorated or even prevented in some patients.

Ethical legal and personal issues

As previously indicated, malingering poses a number of particular challenges to the doctor himself and to the doctor–patient relationship, including amongst others the following.

There is no doubt that seeing substantial numbers of likely ungenuine patients has the potential to affect the practitioner himself. Appropriate 'supervision', as in psychotherapy, may be appropriate.

If the doctor has a treatment responsibility to the patient, he naturally tends to give the patient the benefit of doubt, for example in respect of disability payments. But he will be wise to be cautious: if, for example, he signs a form saying that the patient cannot undertake certain activities of daily living, but without direct observation of that, he could potentially be considered to be an accessory if the claim was subsequently found to be fraudulent.

The use of the word 'malingering' itself in reports must usually be avoided, as it is tantamount to an accusation of a criminal offence such as deception. The medical duty is to present the evidence to the Court, which can then decide the question itself.

Finally, it must not be forgotten that the Data Protection Act gives patients the right to see personal information about themselves such as medical reports. Any suggestion of ungenuine presentation in reports must therefore be well-founded on evidence and properly argued.

Further information

Malleson, A. (2002). *Whiplash and other useful illnesses*. Magill-Queens University Press.

Halligan, P., Bass, C., and Oakley, D. (2003). Willful deception as illness behaviour. In: *Malingering and illness deception* (eds. P. Halligan, C. Bass, and D. Oakley). Oxford University Press.

Vrij, A. (2001). *Detecting lies and deceit*. Wiley.

References

1. Asher, R. (1951). Munchausen's syndrome. *Lancet*, **1**, 339–41.
2. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
3. Turner, M. (2006). Factitious disorders. Reformulating the DSM-IV criteria. *Psychosomatics*, **47**, 23–32.
4. Bass, C. and Halligan, P. (2007). Illness related deception: social or psychiatric problem? *Journal of the Royal Society of Medicine*, **100**, 81–4.
5. Dike, C., Baranoski, M., and Griffith, E. (2005). Pathological lying revisited. *The Journal of the American Academy of Psychiatry and the Law*, **33**, 342–9.
6. King, B. and Ford, C. (1988). Pseudologia fantastica. *Acta Psychiatrica Scandinavica*, **77**, 1–6.
7. Spence, S., Hunter, M., Farrow, T., *et al.* (2004). A cognitive neurobiological account of deception: evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society of London*, **359**, 1755–62.
8. Yang, Y., Raine, A., Lencz, T., *et al.* (2005). Prefrontal white matter in pathological liars. *The British Journal of Psychiatry*, **187**, 320–5.
9. Spence, S. (2005). Letter in *The British Journal of Psychiatry*, **187**, 326–7.
10. Reich, P. and Gottfried, L.A. (1983). Factitious disorders in a teaching hospital. *Annals of Internal Medicine*, **99**, 240–7.
11. Eisendrath, S.J. (1996). Current overview of factitious physical disorders. In *The spectrum of factitious disorders* (eds. M. Feldman and S. Eisendrath), pp. 21–36. American Psychiatric Press, Washington, DC.
12. Krahn, L., Honghzhe, L., and O'Connor, K. (2003). Patients who strive to be ill: factitious disorder with physical symptoms. *The American Journal of Psychiatry*, **160**, 1163–8.
13. Sutherland, A.J. and Rodin, G.M. (1990). Factitious disorders in a general hospital setting: clinical features and review of the literature. *Psychosomatics*, **31**, 392–9.
14. Hengeveld, M.W. (1992). Factitious disorders: what can the psychiatrist do? In *Practical problems in clinical psychiatry* (eds. K. Hawton and P. Cowen), pp. 118–29. Oxford University Press, Oxford.
15. Kerr, B., Foulkes, W., Cade, D., *et al.* (1998). False family history of breast cancer in the family cancer clinic. *European Journal of Surgical Oncology*, **24**, 275–9.
16. Hadeed, V., Trump, D.L., and Mies, C. (1998). Electronic cancer Munchausen syndrome. *Annals of Internal Medicine*, **129**, 73.
17. Feldman, M and Hamilton, J. (2006). Serial factitious disorder and Munchausen by proxy in pregnancy. *International Journal of Clinical Practice*, **60**, 1675–8.
18. Taskaynatan, M.A., Balaban, M., Karlidere, T., *et al.* (2005) Factitious disorders encountered in patients with the diagnosis of reflex sympathetic dystrophy. *Clinical Rheumatology*, **24**, 521–6.
19. Verdugo, R. and Ochoa, J. (2000). Abnormal movements in complex regional pain syndrome: assessment of their nature. *Muscle & Nerve*, **23**, 198–205.
20. Eisendrath, S. and McNeil, D. (2004). Factitious physical disorders, litigation, and mortality. *Psychosomatics*, **45**, 350–3.
21. Peebles, R., Sabella, C., Franco, K., *et al.* (2005). Factitious disorder and malingering in adolescent girls: case series and literature review. *Clinical Pediatrics*, **44**, 237–43.
22. Libow, J. (2000). Child and adolescent illness falsification. *Pediatrics*, **105**, 336–41.
23. Parker, P. (1996). Factitious psychological disorders. In *The spectrum of factitious disorders* (eds. M. Feldman and S. Eisendrath), pp. 37–49. American Psychiatric Press, Washington, DC.
24. DeVaul, R.A. and Faillance, L.A. (1978). Persistent pain and illness insistence. A medical profile of proneness to surgery. *American Journal of Surgery*, **135**, 828–33.
25. Kent, L., Gill, G., and Williams, G. (1994). Mortality and outcome of patients with brittle diabetes and recurrent keto-acidosis. *Lancet*, **344**, 778–81.
26. Barnes, P. and Cheung, K. (1998). Difficult asthma. *British Medical Journal*, **299**, 695–8.
27. Renz, V., Hern, J., Tostevin, T., *et al.* (2000). Functional laryngeal dyskinesia: an important cause of stridor. *The Journal of Laryngology and Otology*, **114**, 790–2.
28. Fink, P. (1992). The use of hospitalisations by persistent somatizing patients. *Psychological Medicine*, **22**, 173–80.
29. Bass, C. and Adshead, G. (2007). Fabrication of illness in children: the psychopathology of abuse. *Advances in Psychiatric treatment*, in press.
30. Eisendrath, S and McNeil, D. (2002). Factitious disorders in civil litigation: twenty cases illustrating the spectrum of abnormal illness-affirming behaviour. *The Journal of the American Academy of Psychiatry and the Law*, **30**, 391–9.

31. Fliege, H., Grimm, A., Eckhardt–Henn, A., *et al.* (2007). Frequency of ICD-10 factitious disorder: survey of senior hospital consultants and physicians in private practice. *Psychosomatics*, **48**, 60–4.
32. Bhugra, D. (1988). Psychiatric Munchausen's syndrome. Literature review with case reports. *Acta Psychiatrica Scandinavica*, **77**, 497–503.
33. Diefenbacher, A. and Heim, S. (1997). Neuropsychiatric aspects in Munchausen syndrome. *General Hospital Psychiatry*, **19**, 281–5.
34. Fehnel, C. and Brewer, E. (2006). Munchausen's syndrome with 20 year follow up. *The American Journal of Psychiatry*, **163**, 547.
35. Kinsella, P. (2001). Factitious disorder: a cognitive behavioural perspective. *Behavioural and Cognitive Psychotherapy*, **29**, 195–202.
36. Eisendrath, S.J. and Feder, A. (1996). Management of factitious disorders. In *The spectrum of factitious disorders* (eds. M. Feldman and S. Eisendrath). American Psychiatric Press, Washington, DC.
37. Teasell, R.W. and Shapiro, A.P. (1994). Strategic–behavioral intervention in the treatment of chronic non-organic motor disorders. *American Journal of Physical Medicine Rehabilitation*, **73**, 44–50.
38. Powell, R. and Boast, N. (1993). The million dollar man: resource implications for chronic Munchausen's syndrome. *The British Journal of Psychiatry*, **162**, 253–6.
39. O'Shea, B., McGennis, A., Cahill, M., *et al.* (1984). Munchausen's syndrome. *British Journal of Hospital Medicine*, **35**, 269–74.
40. Meadow, R. (1977). Munchausen syndrome by proxy: the hinterland of child abuse. *Lancet*, **2**, 343–5.
41. Bools, C., Neale, B., and Meadow, R. (1994). Munchausen syndrome by proxy: a study of psychopathology. *Child Abuse & Neglect*, **18**, 773–84.
42. Elkind, P. (1989). *The death shift: the true story of nurse Genene Jones and the Texas baby murders*. Viking Penguin, New York.
43. Davies, N. (1993). *Murder on ward four*. Chatto and Windus, London.
44. *The Allitt enquiry*. (1994). HMSO, London.
45. Griffin, G. (1996). Assessing dissimulation among social security disability income claimants. *Journal of Consulting and Clinical Psychology*, **64**, 1425–30.
46. Mittenberg, W., Patton, C., Vanyock, E., *et al.* (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, **24**, 1094–102.
47. Larrabee, G. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, **17**, 410–25.
48. Richman, J., Green, P., Gervais, R., *et al.* Objective tests of symptom exaggeration in independent medical examinations. *Journal of Occupational and Environmental Medicine*, **48**, 303–311.
49. Sparr, L.D. and Pankratz, L. (1983). Factitious post traumatic stress disorder. *The American Journal of Psychiatry*, **140**, 1016–9.
50. Rosen, G. (1995). The Aleutian Enterprise sinking and post-traumatic stress disorder: misdiagnosis in clinical and forensic practice. *Professional Psychology: Research and Practice*, **26**, 82–7.
51. Jaffe, M.E. and Sharma, K.K. (1998). Malingering uncommon psychiatric symptoms among defendants charged under California's 'three strikes and you're out' law. *Journal of Forensic Science*, **43**, 549–55.
52. Miller, H. (1961). Accident neurosis. *British Medical Journal*, **1**, 919–25.
53. Mendelson, G. (1995). Compensation neurosis revisited: outcome studies of the effects of litigation. *Journal of Psychosomatic Research*, **39**, 695–706.
54. Grieffenstein, M. and Baker, J. (2005). Miller was (mostly) right: head injury severity inversely related to simulation. *Legal and Criminological Psychology*, **10**, 1–16.
55. Bush S., Ruff, R.M., Tröster, A.I., *et al.* (2005). Symptom validity assessment: Practice issues and medical necessity, NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, **20**, 419–26, NAN position paper.
56. Schmand, B., Lindeboom, J., Schagen, S., *et al.* (1998). Cognitive complaints in patients after whiplash injury: the impact of malingering. *Journal of Neurology, Neurosurgery and Psychiatry*, **64**, 339–43.
57. Word Memory Test (WMT) (Green: wordmemorytest.com)
58. Tombaugh, T.N. (1996). The test of memory malingering. Multi-Health Systems, Toronto, Canada.
59. Smith, G.P. and Burger, G.K. (1997). Detection of malingering: validation of the SIMS. *Journal of the American Academy of Science and the Law*, **25**, 183–9.
60. Goldberg, J.O. and Miller, H.R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *Journal of Clinical Psychology*, **42**, 792–5.
61. Binder, L.M. and Willis, S.C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment*, **3**, 175–81.
62. Rogers, R. (1997). *Clinical assessment of malingering and deception*. Guilford Press, New York.

5.2.10 Neurasthenia

Felice Lieh Mak

Introduction

The term neurasthenia has had a variegated history, and although retained as a diagnostic entity in the ICD-10 it does not appear in the DSM-IV. In cultures where neurasthenia still enjoys popular professional and lay acceptance it has a variety of usages:

- a nosological entity
- an idiom for expressing distress
- a culturally sanctioned illness behaviour
- an explanatory model for a constellation of somatic symptoms
- an euphemism for avoiding the stigma of mental disorder.

Therefore, in diagnosing, understanding, and managing neurasthenia the clinician has to be aware of the context in which the term is used.

Concept and diagnostic entity

The concepts of nervous weakness and asthenia (debility, lack of strength) have existed throughout the history of medicine. Hippocrates described the illness of the Scythians as a general asthenia linked to damage to the genitalia caused by horseback riding. In France, Bouchut (1764) described a syndrome similar to the latter-day neurasthenia, which he called 'neuropathie'. Cullen (1772) conceived muscles and nerves as a unitary nervous force and all diseases as movements against the nature of that nervous force. He coined the word neuroses for this process and postulated that diseases were due to the various alternations of excitement and atony in the nervous system. A few years later, his pupil Brown (1780) elaborated on the hypothesis by dividing diseases into sthenic diseases, which were due to excessive excitement, and asthenic diseases, which were due to deficient excitement. These views on the polarity of the nervous system as a cause of mental illness set the scene for neurasthenia to become a disease entity.

By the beginning of nineteenth century the term neurasthenia was already in use. In 1869, Van Deusen in Holland published a monograph on neurasthenia. This was quickly followed by the publication of a paper, which Beard[1] had presented to the New York Medical Journal Association. Beard based his description of the disorder on a series of 30 cases. In reorganizing the subjective nature of the complaints and the unique clustering of symptoms in each patient, Beard had difficulties in attempting to limit the number of symptoms that constituted the syndrome; he started with 50 symptoms and expanded it to 75 in later publications.

Eventually it became clear that the expanding kaleidoscope of symptoms should be managed in a way that made some sense. Beard approached this problem by organizing the symptoms into subtypes of neurasthenia: cerebrasthenia (cerebral exhaustion) characterized by symptoms that were directly or indirectly connected with the head; myelasthenia (spinal exhaustion) was defined by symptoms related to the involvement of the spinal cord; digestive asthenia was characterized by dyspepsia, constipation, and flatulence. As time went on more subtypes were added by other investigators and specific treatment approaches were developed.

Despite the over inclusiveness of the term, Beard maintained that neurasthenia belonged to one family with a common pathology, prognosis, history, and treatment. As more cases were reported, he felt able to claim that neurasthenia was predominantly an American illness.[2] He attributed the increase in prevalence to the pressures of modern civilization.

Notwithstanding its vagueness, or perhaps because of its vagueness, neurasthenia gained popular acceptance not only by the medical profession but also by the general public. Although by the turn of the century it had become practically a household word, its popularity did not preclude dissent. Most of the criticisms focused on the disorder's over inclusiveness and lack of precision; for instance, Brill called it 'the newest garbage can' in medicine.

The first two decades of the twentieth century witnessed an increasing number of discoveries of more specific causes of disease. This period also saw greater attention being paid to the taxonomy of neuroses. These forces combined to bring about the decline of neurasthenia as a diagnostic entity.

In 1895, Freud published two seminal papers in which he drew up the blueprint for reconfiguring the various neurotic disturbances that were grouped together under the term neurasthenia. In the paper entitled 'On the grounds for detaching a particular syndrome from neurasthenia under the description of 'anxiety neurosis'[3] he questioned the validity of continuing to allow neurasthenia to cover all the symptoms described by Beard. He saw the need to classify different categories of neuroses based on the following:

- collection of symptoms that were more closely related to one another
- common aetiology
- common psychical mechanism.

In the paper 'Obsessions and phobias: their psychical mechanism and their aetiology',[4] Freud removed obsessions and phobias from neurasthenia. As a result of these two papers, neurasthenia ceased to be an amorphous concept and was differentiated into the following categories:

- neurasthenia proper
- anxiety neuroses
- obsessions
- phobias
- pseudoneurasthenias due to cachexia, arteriosclerosis, early stages of the general paralysis of the insane, and psychoses.

Intermittent and periodic types of neurasthenia were to be included under melancholia.

The first list of symptoms Freud proposed for neurasthenia proper included headache, spinal irritation, dyspepsia with flatulence, and constipation. Later, he added sexual weakness and fatigue.

The possibility of including some neurasthenic symptoms under melancholia was mentioned but not expanded on by Freud. This task was taken up by Kraepelin.[5] He distinguished three major types of depression: manic–depressive disorder, involutional melancholia, and a milder form of neurasthenic depression. He asserted that all these types of depression were due to an underlying disordered brain function.

Having been so denuded, the use of the term neurasthenia as a diagnostic entity by the medical professions had declined in the United States by the time of the First World War. The first edition of the DSM-I published in 1952 gave no formal recognition to neurasthenia. Instead, it was replaced by the category of 'Psychophysiological nervous system reaction', the predominant symptom of which was general fatigue. In an effort to make DSM-II congruent with ICD-8, neurasthenia reappeared in American psychiatry as neurasthenic neurosis.

In DSM-III neurasthenia disappeared as an entity and appeared only in the index where readers were asked to refer to 'Dysthymic disorder'. However, unlike the DSM classification, neurasthenia consistently remained a subtype of neurosis throughout the many versions of the ICD. ICD-9 defined neurasthenia as follows.

> A neurotic disorder characterized by fatigue, irritability, headache, depression, insomnia, difficulty in concentration, and lack of capacity for enjoyment (anhedonia). It may follow or accompany an infection or exhaustion or arise from continued emotional stress.

The following categories were included:

- fatigue neurosis
- nervous disability
- psychogenic asthenia
- general fatigue.

Spread to other countries

One of the most fascinating aspects of the history of neurasthenia is its ready acceptance by countries other than the United States where it was originally conceived as a peculiarly American phenomenon. The diagnostic entity took firmer root in some countries than in others. In many countries the concept was indigenized and took on local cultural colour.

The reasons for its spread can be summarized as follows:

- The all-embracing nature of the entity provided a foothold for almost everyone involved.
- The concept provided a blend of scientific theory, thus lending legitimacy to a cluster of symptoms, which are mostly subjective.
- It is considered to be a disease resulting from overwork, which affects the upper social class.

Asia and Australia

In all probability neurasthenia was introduced into China in the 1920s by American psychiatrists and returning Chinese doctors who were trained in the United States. Up to the end of the Second World War, Chinese physicians accepted and used the diagnostic concept of neurosis and neurasthenia from the United States. With the firm establishment of communism in 1949, Pavlovian theory was adopted as the sole model on which Chinese psychiatrists practice, teach, and research.[6] In China, as in the former USSR, neuroses were divided into neurasthenia, psychasthenia, and hysteria. The cause of neurasthenia, as indeed of neuroses, followed the Pavlovian theory of overstrain in the excitation and inhibition processes and mobility of the higher nervous system.

The concept of neurasthenia or *shenjing shuairuo* (nerve weakness), as translated by the Chinese, was not an entirely alien idea. The symptoms associated with neurasthenia (fatigue, loss of memory, poor attention span, headache, tension, insomnia, and all varieties of vague pains) are similar to those in patients suffering from a deficiency in *qi* (vital essence), that is weakness of the kidney, spleen, or heart in traditional Chinese medicine. In addition, the theory of nerve weakness and depletion of nervous energy as causes of neurasthenia fits in with the traditional Chinese medicine concept of organ weakness and *yin–yang* deficiency. Thus in no time at all neurasthenia was incorporated into the body of the practice of traditional Chinese medicine and the vocabulary of the lay public.

In the 1950s, the number of patients suffering from neurasthenia increased enormously. Medical or neurology clinics reported that 80 to 90 per cent of their outpatients were suffering from neurasthenia. It was particularly rampant among the 'brain or mind workers'. The Chinese government regarded it as a serious public health problem, so much so that in its First Five Year Plan (1958–1962) a large-scale national campaign was initiated to eradicate neurasthenia. Research on neurasthenia carried out during this period focused on the role of stress as the external factor, and on heredity and personality as endogenous factors. Treatment included intensive group re-education, herbal medicine, and tranquillizers. Lin[7] postulated that the marked increase in neurasthenia was due to the presence of a deep-seated tension in the revolutionary development of China during the 1950s. Neurasthenia became the vehicle to express political, social, and physical stresses.

About a decade after China's 'open-door policy', an epidemiological survey was conducted in 12 districts in China. The instrument used was the Present State Examination. The results showed that neurasthenia affected 12.59 per cent of persons aged from 15 to 59 years, accounting for 56.7 per cent of all neurotic disorders.[8] In 1982, Kleinman[9] conducted a study of 100 patients diagnosed with neurasthenia in the Psychiatric Outpatient Clinic of the Hunan Medical College. He found that 89 patients satisfied the DSM-III diagnostic criteria for 'Major depressive disorder', 70 per cent of whom responded substantially to antidepressant medication. Despite their improvement, few experienced decreased help-seeking behaviour. This led him to conclude that neurasthenia should be regarded as a special form of somatization related to culturally sanctioned idioms of distress.

In Taiwan, neurasthenia attracted little interest among western-trained doctors. However, it became enormously popular among traditional Chinese doctors, and consequently neurasthenia established itself as a major disease in the minds of the Taiwan public during the 1940s and 1950s.[10]

The mostly British-trained doctors in Hong Kong largely ignored neurasthenia as a diagnostic entity. As in Taiwan, neurasthenia became the domain of traditional Chinese doctors.[11]

In the late nineteenth century psychiatry in Japan was essentially German in orientation. Psychiatrists applied the diagnosis of neurasthenia to patients who presented with weakness, headaches, mental distraction, fatigue, and reduced psychic productivity.[12] The diagnostic entity became a popular term until Morita[13] supplanted it with the term *shinkeishitsu* (nervous or nervous disposition). He described this disorder as basically a psychological reaction to anxiety in predisposed personalities—the personality type being characterized by introversion, perfectionism, hypochondria, hypersensitivity, and self-consciousness. He developed a specific treatment aimed at breaking up the vicious cycle of sensitivity and anxiety, the initial phase of which consisted of isolated bed rest followed by a second phase of work therapy.

Doctors in Malaysia, Singapore, India, Pakistan, Burma, and Sri Lanka are mostly trained in the British tradition. After the First World War neurasthenia lost its popularity in Britain. Standard British textbooks regarded the disorder as rare and outmoded. As a result psychiatrists in these countries tended not to diagnose neurasthenia. However, neurasthenia is used in the Chinese communities where traditional Chinese medicine maintains a stronghold. In India and countries where Ayurvedic medicine is practised, neurasthenia was not added on to the more traditional ways of explaining fatigue, pain, dizziness, and headaches. Instead, concepts such as *dhàtu* loss (loss of semen) and *vàta roga* (wind disease) remained the preferred explanation.

In Australia, Paterson[14] reported that over a 15-year period from 1950 to 1965 neurasthenia was one of the 10 major illness categories reported by a large Sydney-based industry. He claimed that, since he had a fairly representative sample, the 10 categories of illness could very well apply to the rest of Australia.

Europe

From 1880 to 1920 neurasthenia was one of the diseases most frequently discussed. From an 'American nervousness' it rapidly evolved into a western European bourgeois illness. Practically every academic neurologist and psychiatrist wrote a major piece on neurasthenia (see Drinka[15]).

In England, neurasthenia was described in Osler's *The Principles and Practice of Medicine* published in 1900.[16] During the First World War it was a common diagnosis used for invaliding out many soldiers. In order to cope with its diagnosis, treatment, and disposal, the Army instituted a short course of training for medical officers who graduated with the title of 'neurasthenic expert' (see Sims[17]).

Russian psychiatry is largely based on Pavlovian psychophysiological theories. The Pavlovian classification of the principle of neuroses was adopted by all countries that came under the influence of the former USSR. Opinions on the subdivisions of neurasthenia were divided in Russia. One school of thought based its classification on the course of the illness, and the other was based on aetiology.[18] Neurasthenia as a cause of inefficiency and low productivity in the workplace was a recurrent theme in both Russia and Eastern Europe.

Current usage

In ICD-10,[19] neurasthenia is classified as a neurotic disorder in which two main, but overlapping, types of neurasthenia are described:

- the predominant symptom is increased fatigue after mental effort
- predominant feelings of bodily or physical weakness and exhaustion after only minimal efforts.

For a definite diagnosis ICD-10 requires the following:

(a) either persistent and distressing complaints of increased fatigue after mental effort, or persistent and distressing complaints of bodily weakness and exhaustion after minimal effort;

(b) at least two of the following:

- feelings of muscular aches and pains
- dizziness
- tension headaches
- sleep disturbances
- inability to relax
- irritability
- dyspepsia;

(c) any autonomic or depressive symptoms present are not sufficiently persistent and severe to fulfil the criteria for any of the more specific disorders in this classification.

The following are excluded:

- asthenia not otherwise specified
- burn-out
- malaise and fatigue
- postviral fatigue syndrome
- psychasthenia.

DSM-IV does not include neurasthenia as a nosological entity. Instead, it is replaced by 'Undifferentiated somatoform disorder'.

In the third edition of the *Chinese Classification of Mental Disorders*[20]neurasthenia is classified under 'Neurotic disorder'. The criteria for diagnosis have been made more stringent, requiring three symptoms out of five non-hierarchical groups of symptoms, which include weakness, emotionality, excitement, nervous pain, and sleep disturbance. The duration of the symptoms should be at least 3 months. Other psychiatric disorders have to be excluded. Because of the different connotations of fatigue and weakness in Chinese culture, fatigue is not included in the list of symptoms.[21]

Differential diagnosis

Fatigue is a ubiquitous symptom. It can occur in many psychiatric illnesses and in a wide range of physical illnesses. In cultures where the term neurasthenia is loosely used, many of the cases would probably meet the ICD-10 or DSM-IV diagnostic criteria for depressive disorder or anxiety disorder. Physical illness is a common cause of fatigue. In this respect, a detailed history and judicious investigation will be necessary.

Epidemiology

Merikangas and Angst[22] studied a cohort of young adults from a community sample in Zurich, Switzerland, and reported the prevalence of neurasthenia, defined according to the ICD-10 criteria, as 1 per cent across 10 years. The sex ratio across the 10 years of follow-up revealed an equal prevalence among males and females during the initial stages of the study, but females exhibited an l.6-fold greater rate than males during the later stages.

The World Health Organization (WHO) international study[23] of patients with psychological problems seen in primary-care settings reported a prevalence of 1.7 per cent of pure neurasthenia. The prevalence rate increased to 5.4 per cent when the syndrome was diagnosed comorbid with depression or anxiety. The prevalence rate in each centre is shown in Table 5.2.10.1.

The differences in the prevalence rate can be due to many factors, including perception of what health services can treat and the existence of alternative sources of care.

The results of an epidemiological study conducted in 1998 in seven areas in China showed a prevalence rate of 2 per cent.[24]

In a national survey by the Australian Bureau of Statistics Hickie *et al.*[25] reported that 1.5 per cent of the general population met the ICD-10 criteria for neurasthenia in the past year.

All the studies were consistent in demonstrating that the syndrome tended to affect patients below the age of 45 and the absence of significant gender differences.

Aetiology

Although theories abound, the predisposing, precipitating, and perpetuating causes of the syndrome remain unclear.

Table 5.2.10.1 Prevalence of neurasthenic syndrome among patients contacting general health care facilities

Centre	Overall prevalence (%)	Males (%)	Females (%)
Ankara	4.1	1.0	5.6
Athens	4.6	3.3	5.2
Bangalore	2.7	1.7	3.7
Berlin	7.4	4.0	9.7
Mainz	7.7	7.4	8.0
Groningen	10.5	7.1	12.8
Ibadan	1.1	3.4	0.2
Manchester	9.7	6.1	11.3
Nagasaki	3.4	3.8	3.0
Paris	9.3	5.0	14.2
Rio de Janiero	4.5	2.3	5.3
Santiago	10.5	6.4	12.1
Seattle	2.1	2.3	2.0
Shanghai	2.0	1.5	2.2
Verona	2.1	1.8	2.3

Course and prognosis

The 10-year follow-up study conducted by Merikangas and Angst[22] revealed that approximately 50 per cent of patients continued to exhibit symptoms. The WHO study[23] reported that patients with a diagnosis of neurasthenia had, on average, been disabled for 8 to 7 days during the month preceding the examination. Hickie *et al.*[25] confirmed the chronicity of the condition. They reported that 80 per cent of people who met the ICD-10 criteria in the past 12 months were also current cases.

Comorbidity

In the Australian study Hickie *et al.*[25] showed that there was more comorbidity with major depression, panic disorder, and generalized anxiety disorder than could be expected by chance after adjustment for the prevalence of the comorbid disorder and the average level of comorbidity of that disorder.

Treatment

Although there are reports that antidepressants can be effective these are not supported by published data on randomized double-blind controlled trials. Indeed, no such trials have been carried out on any form of psychiatric treatment for neurasthenia. In the absence of such data the clinician will have to rely on the adage of 'When not able to do any good, avoid doing any harm'. Thus aggressive treatment and investigations should be avoided. Patients are best managed in a supportive relationship with due regard given to their psychological and psychosocial needs.

General and non-specific strategies may also be used. These can include regular graded increase in exercise, promotion of sleep hygiene, cognitive techniques to break the cycle of symptoms leading to decreased activities, and improving social support.

Clinicians working in an environment where people with mental illnesses are stigmatized might find it easier to accede to social demands. Clinicians work at two levels in these situations. At one level he or she will have made a diagnosis of a psychiatric disorder and have prescribed the appropriate treatment. At another level the practitioner will be using neurasthenia as a euphemism for mental illness. Therefore, until stigmatization can be reduced or abolished, this unenviable state of affairs will continue.

Complementary medicine

The extracts of leaves from the *Ginkgo biloba* tree contain ginkgoflavone glycosides and trepenoids. *Ginkgo* is widely used as a cognitive enhancer. The roots of the *Panax ginseng* contain several triterpine glycosides, which are believed to have physical performance enhancing properties. This root is used in traditional Chinese medicine to treat a large variety of diseases including neurasthenia. Wesnes *et al.*[26] evaluated the effects of a *Ginkgo biloba/ginseng* combination on 64 healthy volunteers who fulfilled the ICD-10 criteria for neurasthenia. This was a 90-day, double-blind, placebo-controlled, parallel group study. They reported that the combined dose of 120 mg *Ginkgo* extract and 300 mg of *ginseng* extract was significantly better than placebo in reducing the symptoms of neurasthenia by day 90. Adverse effects were nausea and abdominal pain.

The mushroom, *Ganoderma lucidum*, known as lingzhi in China has been widely used to treat cancer, diabetes, and neurasthenia. It is the only known source of ganoderic acid, which has a molecular structure similar to steroid hormones and is a source of biologically active polysaccharide. *Ganoderma* is one of the most highly ranked herbal medicines by Asian people. Tang *et al.*[27] conducted a randomized, double-blind, placebo-controlled parallel study to investigate its efficacy and safety in the treatment of Chinese patients who fulfilled the ICD-10 diagnostic criteria for neurasthenia. Their findings indicated that *Ganoderma* was significantly superior to placebo with respect to the clinical improvement of symptoms in neurasthenic. Adverse effects were mild consisting of nausea, dry mouth, and vomiting.

Complementary medicine is increasingly being used either as an alternative or in addition to conventional psychotropic medications. Refer to Chapter 6.2.9 for further information.

Future directions

As to whether neurasthenia will be replaced by chronic fatigue syndrome (Chapter 5.2.7) or subsumed under somatoform disorders (Chapter 5.2.1), and thus be relegated to a footnote in the history of medicine, or will enjoy resurgence with a new set diagnostic criteria will be determined by future research and clinical data.

Further information

Starcevic, V. (1999). Neurasthenia: cross-cultural and conceptual issues in relation to chronic fatigue syndrome. *General Hospital Psychiatry*, **21**, 249–55.

Tang, W., Gao, Y., Chen, G., *et al.* (2005). A randomized, double-blind and placebo-controlled study of a *Ganoderma lucidum* polysaccharide extract in neurasthenia. *Journal of Medicinal Food*, **8**, 53–8.

Web pages: www.kosmix.com/health Type in neurasthenia to search

References

1. Beard, G.M. (1869). Neurasthenia or nervous exhaustion. *Boston Medical and Surgical Journal*, **3**, 217–20.
2. Beard, G. (1881). *American nervousness: its causes and consequences.* Putnam, New York.
3. Freud, S. (1895). On the grounds for detaching a particular syndrome from neurasthenia under the description of 'anxiety neurosis'. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 1 (ed. J. Strachey), p. 139. Hogarth Press, London.
4. Freud, S. (1895). Obsessions and phobias: their psychical mechanism and their etiology. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 3 (ed. J. Strachey), pp. 74–82. Hogarth Press, London.
5. Kraepelin, E. (1902). *Clinical psychiatry* (trans. A.R. Defendorf), pp. 96–104. Macmillan, New York.
6. Lin, T.S. (1989). Neurasthenia revisited: its place in modern psychiatry. *Culture, Medicine and Psychiatry*, **13**, 105–29.
7. Lin, T.S. (1985). The shaping of Chinese psychiatry in the context of politics and public health. In *Mental health planning for one billion people* (eds. T.S. Lin and L. Eisenberg), pp. 13–14. University of British Columbia Press, Vancouver.
8. Li, C.P. and Zhang, W.H. (1985). Neurasthenia. In *Psychiatry* (ed. Y.C. Shen), pp. 407–12. People's Health Press, Beijing (in Chinese).
9. Kleinman, E.A. (1982). Neurasthenia and depression: a study of somatization and culture in China. *Culture, Medicine and Psychiatry*, **6**, 117–90.

10. Rin, H. and Huang, M.G. (1989). Neurasthenia as a nosological dilemma. *Culture, Medicine and Psychiatry*, **13**, 215–26.
11. Cheung, F. (1989). The indigenization of neurasthenia in Hong Kong. *Culture, Medicine and Psychiatry*, **13**, 227–41.
12. Suzuki, T. (1989). The concept of neurasthenia and its treatment in Japan. *Culture, Medicine and Psychiatry*, **13**, 203–13.
13. Morita, M., Kondo, A. and LeVine, P. (1998). Morita therapy and the true nature of anxiety-based disorders (Shinkeishitsu) translated by Kondo, A. pp. 55–62. SUNY Press, New York..
14. Paterson, G.O. (1969). The economics of illness: an employee sickness study. *The Medical Journal of Australia*, **2**, 249–52.
15. Drinka, G.F. (1984). *The birth of neurosis: myth, malady and the Victorians*. Simon and Schuster, New York.
16. Osler, W. (1900). *The principles and practice of medicine* (7th edn), pp. 166–7. Appleton, New York.
17. Sims, M. (1968). *Guide to psychiatry*, p. 444. Livingstone, Edinburgh.
18. Chatel, J.C. and Peele, R. (1970). A centennial review of neurasthenia. *The American Journal of Psychiatry*, **126**, 1404–13.
19. World Health Organization. (1992). *The ICD–10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. WHO, Geneva.
20. Chinese Psychiatric Society. (2001). *The Chinese classification of mental disorders* (*CCMD-3*). Shandong Publishing House of Science and Technology, Shandong (in English and Chinese).
21. Lee, S. (1996). Cultures in psychiatric nosology: the CCMD–2–R and the International classification of mental disorders. *Culture, Medicine and Psychiatry*, **20**, 421–73.
22. Merikangas, K. and Angst, J. (1994). Neurasthenia in a longitudinal cohort study of young adults. *Psychological Medicine*, **24**, 1013–24.
23. Üstün, T.B. and Sartorius, N. (1995). *Mental illness in general health care: an international study*. Wiley, Chichester.
24. Zhang, W.X., Shen, Y.C., and Li, S.R. (1998). Epidemiological investigations on mental disorders in 7 areas of China. *Chinese Journal of Psychiatry*, **31**, 69–71.
25. Hickie, I., Davenport, I., Issakidis, C., *et al.* (2002). Neurasthenia: prevalence, disability and health care characteristics in the Australian community. *The British Journal of Psychiatry*, **181**, 56–61.
26. Wesnes, K.A., Faleni, R.A., Hefting, N.R., *et al.* (1997). The subjective, cognitive and physical effects of a *Ginkgo biloba*/*Panax ginseng* combination in healthy volunteers with neurasthenic complaints. *Psychopharmacology Bulletin*, **33**, 677–83.
27. Tang, W., Gao, Y., Chen, G., *et al.* (2005). A randomized, double-blind and placebo-controlled study of a *Ganoderma lucidum* polysaccharide extract in neurasthenia. *Journal of Medicinal Food*, **8**, 53–8.

5.3

Medical and surgical conditions and treatments associated with psychiatric disorders

5.3.1 Adjustment to illness and handicap

Allan House

Introduction

Not everybody who develops a serious physical illness will have psychiatric problems as a consequence. To understand why, it is useful to have a model of the normal process of adjustment to stress; psychiatric disorder can then be seen as arising when that process, often called coping, is either maladaptive or is adaptive but only partially successful. This chapter will start with an outline of one theory of stress and coping as it applies to physical illness, followed by a review of disorders of adjustment to illness. A distinction will be drawn between recent-onset illness, which provokes an acute response, and long-standing illness, where the challenge is more often to adjust to chronic disability.

Adjustment to illness and handicap

A number of diseases are reviewed in later chapters, and therefore this chapter will deal with general principles. For more details on particular diseases, the reader should consult specialist textbooks of psychiatry or health psychology.

Illness as a stress

Stress is a word that is used in different ways. Sometimes it refers to an environmental stimulus—a threat or demand from the outside world. This definition lies behind various measures, such as the Social Readjustment Rating Scale[1] or the Bedford College Life Events and Difficulties Schedule[2] which characterize life experiences and produce standardized measures of their severity. According to this view, experiences have properties—as losses, or challenges, or dilemmas—that can be identified by knowing something of the social circumstances of the subject of those experiences but without knowing about the meaning given to them by the person experiencing them.

Another meaning of stress is that it is a bodily state, so that events are only regarded as stressful if they produce changes in the individual. The best-known example of this usage comes from physiology.[3] Stress as a psychological state is also a common lay meaning; when people describe themselves as 'stressed' they are usually referring to a state of tension or autonomic arousal.

Yet another way to understand stress, which is useful in considering physical illness, is that it arises out of an interaction between environmental demands and the resources available to deal with them. This view is articulated in the transactional model of Lazarus and Folkman.[4] According to the theory, when faced with a new experience individuals assesses its likely impact (the primary appraisal) and assess their resources (the secondary appraisal). Stress arises when this double appraisal identifies a mismatch between demands and resources that cannot be narrowed by coping manoeuvres.

(a) Illness as a demand or threat (the primary appraisal)

There are a number of characteristics of an experience that increase the chances of it being appraised as threatening. These include immediacy, ambiguity, uncontrollability, or undesirability. The probability that many people will share an interpretation of a particular episode explains the similarity of people's responses to certain illnesses. The possibility of individual, even idiosyncratic, interpretations can explain sharp differences between people with apparently the same disorder.

A useful way to construe individual appraisals of illness is outlined by Leventhal *et al.*[5] in their theory of internal illness representations. The common elements of the illness representation can be identified from a simple self-report questionnaire[6].

Illness beliefs cannot be assumed solely on the basis of the illness from which a person is suffering, or from his or her social context.[7] Individuals may hold unpredictable beliefs—that an illness is inherited from a family member, or that it is a punishment for a misdemeanour, or that it may be curable by adopting an unusual diet. For some, the representation of illness overlaps with the representation of self, so that sufferers see themselves as living their illness rather than suffering from it.[8] (see Box 5.3.1.1)

The characteristics of a particular disease are not the only component of the illness that can make it threatening. Illness occurs in a social and interpersonal context, and while the responses of other people may be helpful, they may in some cases contribute to the demands of the situation. For example, a partner may withdraw or become depressed, or family members may become intrusive or overcontrolling. Being ill confers a special status, the so-called sick role, but it is a status acquired at a cost in the loss of independence and certain rights. While disability may arise largely from the impairments caused by a disease, much handicap is socially determined.

(b) Resources for responding to illness (the secondary appraisal)

The focus of secondary appraisal is twofold: the person's personal resources, and the resources external to them, mainly in the immediate social network.

Personal resources may be defined in a number of ways, for example cognitive attributes, personal characteristics, or personality traits.

The other resource for the individual is social support. There are many approaches to understanding support, but a useful one[9] is to regard it as having four components:

1 emotional support, conveying a sense of being cared about or loved

2 esteem support, conveying a sense of being valued or respected

3 instrumental support, conveying practical help

4 informational support, conveying knowledge relevant to tackling the problem

Box 5.3.1.1 Components of the illness representation

- identity (label and associated symptoms)
- causal ideas
- consequences (severity and likely impact)
- time-line (natural history)
- curability or controllability

The family's reaction to illness has an important impact on the type of support available. If they are rejecting, intolerant of dependence, or unsympathetic to the needs of the patient—for example, to change their diet, or stop smoking, or take more (or less) exercise—then they may offer too little support. On the other hand, they may be overprotective, refusing to allow the patient a reasonable degree of autonomy and discouraging active coping. Sometimes, members of a family will hold different views about the nature of an illness, leading to conflict, which is not always revealed to doctors. More often, they share views. If such views are inaccurate (so-called family myths) and yet strongly held, then they can be a powerful barrier to the patient accepting medical advice. It is a common observation that patients with chronic illness who are depressed often have a carer who is depressed, and this tendency to share (often dysfunctional) beliefs and coping styles is one reason for that.

(c) Coping with illness

Coping refers to efforts to reduce the gap between demands and resources. Coping is described according to its aims, the techniques used to achieve those aims, and according to the overall coping style adopted.

The ***aims*** of coping are either problem focused, designed to modify the demands of the situation, or emotion focused, designed to modify how one feels about a situation.[4] Emotion focused coping generally works well but only transiently. It is best reserved for brief stresses, such as unpleasant medical procedures, or for situations in which nothing can realistically be done to modify the stress.

The ***techniques*** for coping serve to mobilize available resources. Vocabularies differ for describing them. Cognitive coping techniques include information seeking, downplaying, or adopting a defiant or overoptimistic attitude. In psychodynamic terms, the two most commonly used techniques are probably denial and regression. In common usage, the techniques referred to by these vocabularies overlap. Behavioural coping may involve changing ones lifestyle, such as exercising more or excessive drinking of alcohol. Social coping is a particular form of behavioural coping, and may involve increasing contacts or accepting help from professional agencies. In chronic illness, successful coping may be accompanied by a slower process of reappraisal—in which the patient comes to a different understanding of the illness, from that apparent at initial diagnosis—through for example *benefit-finding* and *downward comparison* (with others who have worse disability, pain, or whatever).

Coping ***styles*** are more general approaches to coping. Two contrasted styles are active/engaged (sometimes called 'approach') coping and passive/disengaged (sometimes called 'avoid') coping.[10] While it is appealing to characterize people as having a particular coping style, and while it is possible to think of typical examples from personal experience, in fact most people do not have a sufficiently unchangeable repertoire of coping techniques to merit the label of a style.

Adjustment disorders

(a) Definition and classification

The emphasis in ICD-10[11] is on emotional disturbance as the characteristic feature of adjustment disorders—some disturbance of behaviour is acknowledged, particularly in adolescence.

However, it is common to encounter cognitive or behavioural changes that interfere with social functioning and quality of life, and yet which are not attributable to the consequences of mood disorder. DSM-IV[12] acknowledges this possibility more directly, including a category of 'Adjustment disorder, unspecified', which covers 'maladaptive reactions (e.g. physical complaints, social withdrawal, or work or academic inhibition)'.

Examples of cognitive problems are extreme helplessness, denial of the existence of illness, or of the handicap associated with it. Behavioural problems may include marked social withdrawal or lack of self-care, or irrational non-adherence to treatment. Emotional problems are typically thought of as anxiety or depression, but irritability is also common.

(b) Diagnosis and differential diagnosis

The diagnostic features of adjustment disorders are relatively non-specific, comprising mood symptoms and behaviour disturbances, which do not meet the criteria for a diagnosis of another disorder, and yet which are sufficient to amount to a mental disorder. The two main diagnostic questions are as follows.

- Does the patient have a diagnosable mental disorder?
- If there is a mental disorder, should it be given another more specific label than 'adjustment disorder'?

What distinguishes normal adjustment from a disorder? The first criterion is whether the symptoms are persisting beyond the time when they might be attributable to the stressor. This judgement is relatively straightforward when the stressor is a single event. However, if illness is more persistent or intermittent—such as cancer followed by intensive treatment, or multiple sclerosis—then it is less easy to judge.

The second criterion is whether the response is causing avoidable social dysfunction. For example, in many cultures illness is followed by a period of convalescence, during which activity is reduced and a return to full social responsibilities is deferred. This may be a healthy avoidance of activity, if it allows full recovery from illness, but prolonged avoidance of activity may lead to secondary physical problems as well as social isolation and loss of role.

When adjustment disorders are associated with chronic illness and handicap, the duration criterion cannot apply. An individual may present symptoms because his or her response is outside the culturally acceptable range; for example, he or she may be too demanding or uncooperative, or too passive and dependent. It is unwise to regard a presentation as disordered simply on these grounds. The best indicator is whether the individual is achieving the highest level of function and the lowest level of distress of which they are capable under the circumstances. This means that each person must be diagnosed according to his or her own context, and that a standardized set of criteria cannot be applied.

The differentiation of adjustment disorders from other psychiatric disorders is more straightforward, and depends on the presence or absence of key symptoms. The main conditions found in association with physical illness are depressive disorders, anxiety disorders, and occasionally post-traumatic stress disorder.

(c) Epidemiology

Little is known about the epidemiology of adjustment disorders other than those involving mood disturbance, because of the absence of standardized diagnostic criteria.

Psychiatric symptoms are distributed in the general population, with a positive skew to the distribution. In the physically ill, the same pattern of distribution is seen, but the curve is shifted to the right. The increase in psychiatric symptoms is contributed to by a general increase in all the common symptoms. The usual way to identify cases is to select those who cross an accepted threshold for symptom levels—as determined, for example, by one of the standardized self-report questionnaires—and then to apply diagnostic criteria. Adopting this approach, rates of diagnosable mood disorder among the physically ill are about double what they are in the general population. That is, 30 to 50 per cent (depending on the population studied and the diagnostic criteria employed) of the physically ill have a mental disorder. Approximately two-thirds of these cases are adjustment disorders, the rest meeting criteria for another disorder (usually depressive).

The elderly report lower rates of psychiatric disturbance. This may be a cultural effect, with the elderly disposed to report fewer symptoms of distress as a result of stoicism learned through experience of adversity earlier in life. Alternatively, the elderly may genuinely respond differently to physical illness.

Mood symptoms and adjustment disorders are commoner in response to acute illness than they are in chronic illness.

(d) Aetiology

There are several reasons why coping might fail.

First, demands may be overwhelming. The news that one has a terminal illness takes time to assimilate—to understand all its meanings, grasp all the threats and losses involved. While that process of appraisal is going on, it is difficult to marshal resources and use them effectively. This explains, in part, why mood disorder is more commonly associated with acute than chronic illness.

Second, resources may be inadequate or missing. One problem associated with physical illness is that it may impair personal resources as a primary effect of the disease process—most importantly when the illness has effects on the central nervous system by virtue of the direct involvement of the brain or through the neurological effects of systemic disturbance.

Third, coping responses may be ineffective. There are few rules about what makes effective coping. In general, a broad and flexible repertoire is desirable, with a strong element of active problem-focused techniques. However not all illnesses, nor all aspects of a particular illness, are likely to be amenable to problem-focused coping. Probably the most effective coping is matched to the situation. That is, the coping matches the demands, so that heavy reliance is not placed on problem-focused coping when little in the situation can change, nor excessive use made of emotion-focused coping when active involvement in illness management is needed.

A common problem of failure to match coping to the situation is found in patients with chronic illness, who are responding to their circumstances as if they none the less have an acute illness. In acute illness, problem-focused coping often involves seeking reversal or even cure of the illness process, while emotion-focused coping involves dealing with the anxiety of uncertainty, or grieving if the prognosis is clearly poor. On the other hand, in chronic illness, problem-focused coping involves symptom management and maximizing function, while emotion-focused coping requires a degree of acceptance.

It is not easy to predict who will develop an adjustment disorder. Certainly the risk is not strongly linked to physical diagnosis, or

within a particular diagnostic group to physical disability. The most robust finding is that a previous history of psychiatric problems increases the risk of psychiatric problems associated with physical illness.

(e) Course and prognosis

By definition, adjustment disorders arise shortly after diagnosis. In practice, there is variation; some people respond immediately and develop symptoms within days, while others develop symptoms weeks or even months after diagnosis. The losses associated with illness may only become apparent when a person leaves hospital and faces functional impairment at home. Carers and others in the social network respond differently to acute and chronic illness, and it may take time for that to become clear. The greater the delay from the onset of illness to the emergence of symptoms, the harder it is to make a diagnosis of adjustment disorder. In clinical practice, it is reasonable to set an upper limit of a year.

Most adjustment disorders provoked by a newly onset illness, resolve within weeks. Slower recovery takes place over 12 or 18 months. If recovery has not occurred by then, the patient has usually developed another mental disorder, such as a depressive disorder. Accurate data are few, but probably no more than 10 per cent of patients develop a prolonged adjustment disorder.

The psychiatric symptoms of adjustment disorder impair quality of life, so much so that all standardized quality-of-life measures include mood symptoms in their profile. Psychiatric morbidity associated with physical illness is also a risk factor for self-harm and for completed suicide. Adjustment disorders are likely to have an effect on the outcomes of treatment for physical disease.[13] Health service costs are greater for patients with physical illness and psychiatric co-morbidity; lengths of stay are longer for hospital inpatients; the functional outcomes of rehabilitation may be poorer, and there is some evidence that there may also be an increased mortality. The mechanism for these effects may be broadly behavioural or physiological. Examples of the former are increased rates of smoking, lack of exercise, and poor adherence to treatment regimes among people with mental disorders. Examples of the latter include activation of the hypothalamic–pituitary–adrenal axis or increased cytokine production associated with chronic emotional disorder.

(f) Treatment

(i) Drug treatments

Antidepressants can be effective in the presence of physical illness.[14] There is no good evidence to support claims for a great superiority in efficacy of serotonin-reuptake inhibitors. Although their long-term tolerability may be greater than older drugs in patients with physical disease, they are not without toxicity—for example they have been associated with increased falls and gastrointestinal haemorrhage. Tricyclic antidepressants have advantages in treating patients with insomnia or chronic pain. Cost differences are substantial.

(ii) Psychological treatments

A number of brief psychological therapies have been shown to be effective in treating depression; namely cognitive-behaviour therapy, problem-solving therapy, interpersonal therapy, and brief dynamic therapy. Such therapies may also be effective in treating adjustment disorder in the physically ill,[15,16] although they may have disappointingly weak effects upon physical outcomes. Therapy may need to be modified to allow for fatigue or concentration problems, and sessions need to be arranged flexibly to accommodate hospital appointments and other treatment needs.

(g) Management

(i) Identifying cases

A major difficulty in delivering treatment to people with adjustment disorders is the difficulty in identifying cases. There are several self-report questionnaires, which may be used to screen for patients with mood disorder. In certain settings such questionnaires can be delivered routinely to all patients, for example by means of computers with touch screen technology, and they are useful for alerting staff to the presence of mood symptoms—but their positive predictive value is too low to allow for accurate use in identifying those who need referral to specialist services. Their use is also difficult to integrate into routine clinical practice, and response rates outside research studies are usually low.[17] Their use is best restricted to specialist services where the clinical staff are clear about what response they will make to a high score, since this is where there is some evidence for the benefits of case-finding. There are no useful standardized instruments for the detection of other problems with adjustment.

Instead, clinicians should be encouraged to consider the possibility of psychiatric disorder when there is a gap between impairment and handicap so that the patient is doing worse in rehabilitation than the severity of their disease would suggest they should be, when there are multiple complaints that are difficult to explain, or when multiple drug treatments are being administered without conspicuous benefit. The clinical interview is the mainstay of diagnosis.

There are a number of common reasons for failing to recognize adjustment disorders. First, the questions simply are not asked, or attempts by the patient to introduce the topic of psychological problems are blocked or sidestepped. Second, questions may be asked, but in circumstances where it is difficult for the patient to answer honestly—when there is no privacy, or the person asking is obviously too busy to listen to any but a conventional answer. Third, expressions of distress may be normalized, and thus dismissed: 'Of course it's natural you will feel like that' means to the patient 'So please don't mention it again'.

(ii) Broadening the repertoire of psychological responses

No single intervention is going to be effective for all patients with psychological problems arising from difficulties with adjustment to illness. Realistic management therefore involves offering what has been called a 'menu of interventions'.[18] A currently favoured model is a so-called ***stepped care*** model in which intervention is organized in a hierarchy according to intensity of treatment and the expertize needed to deliver it e.g. see Box 5.3.1.2. While unfocused 'support' is of limited value (because it does not encourage active secondary appraisal and experiments with different coping strategies) there are now many brief and flexible psychological therapies available, some of which may be deliverable by staff in the primary care or physical healthcare service (see Box 5.3.1.3). In ***collaborative care*** such first line psychological treatments, along with medication management and simple social care, are delivered by non-mental health staff and monitored by a case manager, with a mental health professional providing supervision and back-up consultation. Successful trials have been conducted in (for example) heat disease[19] and diabetes.[20]

Box 5.3.1.2 Stepped care in the treatment of depression (NICE clinical guideline 2004)

	Who is responsible for care?	What is the focus?	What do they do?
Step 5	Inpatient care, crisis teams	Risk to life, severe neglect	Medication, combined treatments, ECT
Step 4	Mental health specialists including crisis teams	Treatment-resistant, recurrent, atypical and psychotic depression, and those at significant risk	Medication, complex psychological interventions, combined treatments
Step 3	Primary care team, primary care mental health worker	Moderate or severe depression	Medication, psychological interventions, social support
Step 2	Primary care team, primary care mental health worker	Mild depression	Watchful waiting, guided self-help, computerized CBT, exercise brief psychological interventions
Step 1	GP, practice nurse	Recognition	Assessment

National Institute for Health and Clinical Excellence (NICE) (2005). CG22 Anxiety: quick reference guide. London: NICE. Available from http://www.nice.org.uk/nicemedia/pdf/CGO22quickrefguideamended.pdf. Reproduced with permission.

Box 5.3.1.3 Brief psychological treatment of use in the medically ill

- **Motivational interviewing** is an approach developed to encourage people to attempt change in addictive behaviours. It may be useful in engaging people in demanding treatments, or in improving adherence to treatment regimes.
- **Graded activity** has been used to treat negative symptoms in mental illness like schizophrenia or depressive disorder. It is effective in improving function in chronic fatigue syndrome, and is worth using in other conditions where inactivity and passivity is out of proportion to physical disability.
- **Anger management** is a modification of cognitive-behaviour therapy, which may be useful where irritability or aggressive behaviour is complicating adjustment.
- **Interpersonal therapy**[21] was initially developed for the treatment of depression, but it has obvious applications in the field of physical illness. In the terminology of interpersonal therapy, illness represents a role transition, and the focus in therapy is therefore on negotiating that transition with key others in the patient's life.
- **Family therapy and couples therapy** are rarely considered (or available) for adults with physical illness, and yet many of the external resources needed for coping are in the family.

For more severe or persistent problems, referral to specialist services is appropriate—ideally liaison services that operate in the primary or secondary care settings where patients receive their main care. Psychiatric treatment of the physically ill, especially in hospital, requires a number of modifications to routine clinical practice, which are sometimes overlooked.

First, an extra effort has to be made to meet the family and carers. They may be reluctant to attend if there is hostility in the family, or if missed time from work is creating financial pressures, but failure to interview others makes it near impossible to come to a full and accurate formulation of the problem.

Second, personal contact with the referrer is highly desirable. The 'real' question may not be that posed in the referral, and can only be identified by probing. Advice is much more likely to be followed if it is delivered face to face, and followed up with a later visit to check on compliance! This direct contact with non-psychiatric colleagues is one of the defining characteristics of liaison psychiatry, and its importance cannot be overemphasized.

Third, it must be recognized that the course of psychiatric treatment needs to be modified. Appointments will be missed, or interrupted, by the demands of physical treatment. And psychological issues may well not be resolved by a single clinical encounter; a relapse of illness may provoke a further episode with new features, and patients often have to return repeatedly to work through themes in therapy, as they are re-challenged with new physical problems.

(h) Prevention

There are two broad approaches to prevention, namely education and support.

Education and the provision of information and advice about the illness and its management is desirable as an informed patient is more likely to be an effective partner in treatment, and because it is popular with patients. Disappointingly, however, it is not an effective means of preventing psychiatric problems. This is probably because, while it facilitates primary appraisal, it does nothing to facilitate secondary appraisal or the use of effective coping strategies.

Provision of support is also popular. It takes a number of forms, including self-help groups, volunteer visiting, and professional support workers—usually with knowledge of a particular disease such as AIDS or a stroke. Again, there is little evidence that it prevents psychiatric problems. Perhaps this is because it usually provides emotional support; which in itself may be worthwhile but which is insufficient if not combined with a more problem-focused approach.

In conclusion, there are no clear indications that we can prevent the development of adjustment disorders. The mainstay of current management is therefore to identify existing cases and to offer specialist care to those who are most symptomatic or handicapped, and to those who are not improving spontaneously.

Adjustment to terminal illness: care of the dying

Adjustment to terminal illness has much in common with adjustment to other severe illness, and is not specifically the province of psychiatrists. For a detailed discussion of the care of the dying, the reader is referred to a more specialized text.[22] Here we will discuss two issues that are commonly presented to the psychiatrist in this setting: the diagnosis of depression and other adjustment disorders, and the issue of suicide.

Diagnosis of depression and adjustment disorders in the terminally ill

As with physical illness, somatic complaints are common in the dying and thus, individually, lose their diagnostic or predictive usefulness in the major depressive syndrome. Even psychologically, a degree of hopelessness may be appropriate. Anxiety is a common symptom in the dying, but it is not necessarily pathological. Like depression, it may result from physical disability, uncontrolled pain, or pre-existing anxiety disorders. In these circumstances a more detailed examination of the attitudes of the patient is necessary. Pervasive global hopelessness, feelings that life has had no meaning, strong feelings of guilt or punishment, and suicidal thoughts are pointers towards depressive illness in the terminally ill.

Assessing suicidal thoughts in the terminally ill

Several studies have shown that the prevalence of suicidal thoughts among patients with terminal cancer is less than 10 per cent.[23] However, this contradicts the clinical impression that most patients admit to either suicidal thoughts or thoughts of assisted suicide as an escape from the imaged consequences of losing control. In some patients, having a belief in a 'way out' can be positive in offering a sense of control.

Completed suicide is an important complication in patients with a terminal illness. General predictors of suicide apply, with the addition of severity of functional impairment, isolation, and delirium. The two most important factors to watch out for are uncontrolled pain and depression. These two factors greatly increase suicide risk but are nevertheless treatable in the terminal illness setting.[24]

People who express a desire to die are nearly always ambivalent. The expression of suicidal ideas should never be accepted as rational without a searching enquiry for evidence of subtle external pressures, fear of terminal symptoms or of being a burden on others, and treatable depression.

Pharmacological management of psychiatric disorders in the dying

Alleviation of distress rather than cure is the guiding principle of the management of the terminally ill. Caution is needed in the selection and prescription of psychotropic drugs.

Anxiolytic medication is usually well tolerated and the concern over dependence and tolerance is less of an issue. Terminal metabolite benzodiazepines, such as lorazepam and oxazapam, can provide symptomatic relief for a variety of conditions even in patients with hepatic impairment. Occasionally, opiates are used in this role where first-line treatments are unsuccessful.

Antidepressant use in the dying may be more problematic due to the adverse effects of sedation, seizures, hypotension, and constipation and urinary retention. For this reason, the choice of drug needs to be individually tailored. Dosage has to be carefully adjusted, beginning at low doses and increasing gradually. The use of psychostimulants (dextroamphetamine, methylphenidate) is worth considering. These drugs are used infrequently in general psychiatry because of the risks of dependence, but in the terminally ill they may have advantageous 'energizing' properties, including increased energy, improved concentration, increased appetite, and possibly, a faster onset of action.

Further information

Sharpe, L. and Curran, L. (2006). Understanding the process of adjustment to illness. *Social Science & Medicine*, **62**, 1153–66.

Fletcher, J., Bower, P., Gask, L., *et al.* (2006). *Primary care services for depression: a guide to best practice care services improvement partnership*. National Institute for mental Health in England, Department of Health, London.

References

1. Holmes, T.H. and Rahe, R.H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, **11**, 213–18.
2. Brown, G. and Harris, T. (1978). *Social origins of depression*. Tavistock Publications, London.
3. Selye, H. (1956). *The stress of life*. McGraw–Hill, New York.
4. Lazarus, R.S. and Folkman, S. (1984). *Stress, appraisal and coping*. Springer, New York.
5. Leventhal, H., Nerenz, D., and Steele, D. (1984). Illness representations and coping with health threats. In *Handbook of psychology and health* (4th edn) (eds. A. Baum, S.E. Taylor, and J.E. Singer), pp. 219–52. Erlbaum, Hillsdale, NJ.
6. Weinman, J., Petrie, K., Moss–Morris, R., *et al.* (1996). The illness perception questionnaire: a new method for assessing the cognitive representations of illness. *Psychology and Health*, **11**, 431–45.
7. Petrie, K. and Weinman, J. (eds.) (1997). *Perceptions of health and illness*. Harwood, Amsterdam.
8. Kleinman, A. (1988). *The illness narratives: suffering, healing and the human condition*. Basic Books, New York.
9. Cohen, S. and Wills, T.A. (1985). Stress, social support and the buffering hypothesis. *Psychological Bulletin*, **98**, 310–57.
10. Carver, C.S., Weintraub, J.K., and Scheier, M.F. (1989). Assessing coping strategies: a theoretically based approach. *Journal of Personality and Social Psychology*, **56**, 267–83.
11. World Health Organization. (1992). *International statistical classification of diseases and related health problems,* 10th revision. WHO, Geneva.
12. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
13. Saravay, S.M. and Lavin, M. (1994). Psychiatric comorbidity and length of stay in the general hospital–a critical review of outcome studies. *Psychosomatics*, **35**, 233–52.
14. Gill, D. and Hatcher, S. (2000). Antidepressants for depression in medical illness. *Cochrane Library*, **4**.
15. MacHale, S. (2002). Managing depression in physical illness. *Advances in Psychiatric Treatment*, **8**, 297—305.
16. Winkley, K., Landau, S., Eisler, I., *et al.* (2006). Psychological interventions to improve glycaemic control in patients with Type I diabetes: systematic review and meta-analysis of randomized controlled trials. *British Medical Journal*, **333**, 65–8.
17. House, A. (1988). Mood disorders in the physically ill: problems of definition and measurement. *Journal of Psychosomatic Research*, **32**, 345–53.
18. Goodheart, C.D. and Lansing, M.H. (1997). *Treating people with chronic disease: a psychological guide*. American Psychological Association, Washington, DC.
19. Schrader, G., Cheok, F., Holdacre, *et al.* (2003). Effect of psychiatry liaison with general practitioners on depression severity in recently hospitalized cardiac patients: a randomized controlled trial. *JAMA: The Journal of the American Medical Association*, **288**, 2836–45.
20. Williams, J.W., Katon, W., Lin, E.H.B., *et al.* (2004). The effectiveness of depression care management on diabetes-related outcomes in older patients. *Annals of Internal Medicine*, **140**(12), 1015—24.
21. Klerman, G.L. and Weissman, M.M. (1993). *New applications of interpersonal psychotherapy*. American Psychiatric Press, Washington, DC.

22. Wiener, I., Breitbart, W., and Holland, J. (1996). Psychiatric issues in the care of dying patients. In *Textbook of consultation–liaison psychiatry* (eds. J.R. Rundell and M.G. Wise), pp. 804–31. American Psychiatric Press, Washington, DC.
23. Brown, J.H., Henteleff, P., Baratat, S., *et al.* (1986). Is it normal for terminally ill patients to desire death? *The American Journal of Psychiatry*, **143**, 208–11.
24. Bolund, C. (1985). Suicide and cancer II: medical and care factors in suicide by cancer patients in Sweden. *Journal of Psychosocial Oncology*, **3**, 17–30.

5.3.2 Psychiatric aspects of neurological disease

Maria A. Ron

Psychiatric abnormalities are an integral part of neurological disease and their study can improve our understanding of the neural basis of psychiatric illness. This chapter deals with common neurological diseases where psychiatric symptoms are prominent.

Stroke

Stroke is defined as the sudden loss of blood supply to an area of the brain resulting in permanent tissue damage and is the commonest neurological disorder. The incidence of stroke for those aged between 35 and 65 is between 90 and 330 per 100 000. It is commoner in men and the incidence increases with advancing age. Ischaemic stroke is commoner than haemorrhagic stroke and accounts for 80 to 85 per cent of all cases.

Depression

Its prevalence is around 30 per cent in the first few weeks after a stroke—two-thirds of the patients fit the criteria for major depression and the rest for minor depression. Survivors remain at an elevated risk for depression for many years.

(a) Clinical features

- Diurnal variation of mood, weight loss, anergia, insomnia, and loss of libido are prominent in the early stages.
- Anhedonia, suicidal ideation, loss of self-esteem, and feelings of guilt become evident later.
- Irritability and aggressive behaviour are common, especially in those with cognitive impairment.[1]
- The onset of depression may occur acutely in the early post-stroke period or be delayed for 6 months or more.

(b) Factors associated with post-stroke depression

Early after stroke, anterior left-sided lesions involving the cortex and subcortical regions, especially the basal ganglia, and right posterior lesions are more frequently associated with depression.[2] In time these associations become less marked.[3] Cognitive impairment is closely associated with depression early after stroke,[4] and is present in 70 per cent of those with major and 43 per cent of those with minor depression. Old age, a past or family history of depression, and negative life events in the preceding 6 months substantially increase the incidence of post-stroke depression.

Disruption of fronto-subcortical circuits, directly or as a distant effect of stroke, plays a central role in the causation of depression. Decreased metabolism in orbitofrontal, anterior cingulate, and inferior temporal regions has been reported using **PET**.[5] Serotonergic mechanisms have been implicated, with a reduction of 5-hydroxytryptamine-2 receptor binding in the temporal cortex, especially in left hemisphere stroke.

(c) Course and prognosis

The average duration of major depression is around 1 year, with spontaneous remission in many patients. Symptoms of minor depression may persist for 2 years or more. The presence of depression and other psychiatric diagnosis in the first 3 years after a stroke substantially increases the risk of death after controlling for cardiovascular and other risk factors.

Anxiety

- About a quarter of patients fulfil criteria for generalized anxiety disorder during the acute post-stroke phase.
- Rates are lower in community studies (5 per cent)[6] and 1 or 2 years after stroke (4 to 18 per cent). Half of those with anxiety disorder also satisfy criteria for depression.
- Right-sided subcortical lesions may be more common in anxiety disorder, while left-sided pathology, usually involving cortical regions, is more likely when anxiety and depression coexist.[2] Serotonergic abnormalities are also likely to be relevant.

(a) Course and prognosis

Anxiety in the acute post-stroke phase is associated with high mortality rates, comparable to those in depressed patients.

Emotionalism (abnormal crying or laughing)

Emotionalism is usually mood-congruent and is triggered by sad or emotional events. Most patients have some degree of voluntary control.

Emotionalism occurs in a quarter of patients during the first year post-stroke, with a peak in the first month and decreasing gradually thereafter.[7] It is associated with the severity of depression and with left-anterior lesions disrupting serotonergic pathways.

Management

The first step in treating the psychiatric manifestations of stroke is for these to be recognized, and patients need to be routinely assessed for the presence of psychiatric symptoms.

Treatment

Double-blind placebo-controlled trials using nortriptyline, trazodone, and selective serotonin reuptake inhibitors (**SSRI**s) have shown these drugs to be effective.[4] Improvement in depression also results in lasting improvement in cognition and physical activity. Treatment within 3 months of stroke may be followed by the best outcome.

The treatment of anxiety symptoms has been less well documented. Short-acting benzodiazepines, buspirone, and SSRIs are the main pharmacological approaches. The usefulness of psychotherapeutic or behavioural interventions has not been established

and may depend on the severity of symptoms and cognitive impairment.

Emotionalism responds well to treatment with SSRIs and tricyclic antidepressants, even in those without associated depression.

Parkinson's disease

The neurological and cognitive features of Parkinson's disease are dealt with in Chapter 4.1.6 and only commonly encountered psychiatric symptoms will be considered here.

Depression

Its overall prevalence is approximately 40 per cent.[8] Depressive symptoms are more common early in the disease (50 per cent) and in those with onset before the age of 55. For many people, adaptation to the disease results in a return to normal mood. Depression becomes more frequent again in the advanced stages of the disease, particularly in those with rapidly progressive disability.

Major depression is commoner in those with akinetic Parkinson's disease (38 per cent) than in those with classical forms of the disease (15 per cent), but dysthymia is equally common in both. Depression, severe in some patients, has also been reported in the first few weeks after bilateral subthalamic stimulation, a successful treatment for the motor symptoms of Parkinson's disease,[9] but mood changes were less commonly observed a year after surgery.

(a) Clinical features

- Anxiety, agitation, and depressed mood are prominent.
- Depressive symptoms are not closely associated with the severity of motor signs, but may be more severe during 'off' periods and are commoner in those with cognitive impairment.
- Early awakening, motor retardation, and apathy in the absence of mood abnormalities may not be indicative of depression.
- Major depression is associated with functional deterioration at 1-year follow-up.[10]
- Three-quarters of patients with depression also fulfil criteria for anxiety disorder, but only 10 per cent have symptoms of anxiety in isolation.[11]

(b) Mechanisms underlying depression

Studies using PET have suggested that depression and anxiety in Parkinson's disease are associated with a specific loss of dopamine and noradrenaline innervation in the limbic system.[12] Postmortem studies have also described loss of dopaminergic neurones in the ventral tegmental area.[13] Degeneration of these neurones leads to dysfunction of the orbitofrontal cortex with secondary effects on the serotonergic cell bodies of the dorsal raphe.[5]

(c) Management

Optimal control of neurological symptoms may lead to improvement in depression and should be a management aim. Repeated assessment may be needed to differentiate features of the disease, such as apathy, from the symptoms of depression.

(d) Treatment

There are few studies describing the treatment of depression in Parkinson's disease. When antidepressants are clinically indicated because of the severity or persistence of symptoms, the antidepressant profile of side-effects should be considered. Antidepressants with strong anticholinergic effects, such as amitriptyline, may increase cognitive impairment and SSRIs may be preferable. Electroconvulsive treatment is also effective for Parkinson's disease patients with depression and may also transiently improve motor symptoms.

Psychotic symptoms

(a) Clinical features

- Up to 40 per cent of patients on long-term treatment experience visual hallucinations.
- Long duration of illness, age, cognitive impairment, and depression are associated with visual hallucinations.
- Visual hallucinations in clear consciousness are usually fully formed images of people or animals, non-threatening, fleeting, and stereotyped. They are recurrent and tend to occur at night, more commonly in the 'on' periods.
- Sleep disturbances (fragmented sleep, alteration of sleep rhythms, and vivid dreams) often precede daytime hallucinations and may be part of a continuum.
- Delusions are less frequent than hallucinations but are more stressful and difficult to manage. They are usually paranoid, with conspiracy and infidelity themes.
- Severe psychosis is associated with institutional placement, progressive dementia, and increased risk of death (over a quarter of patients within 2 years).

(b) Mechanisms

Psychotic symptoms may represent intrusion of REM sleep imagery into wakefulness. They may be more frequent in those receiving anticholinergics and dopamine agonists, but there is no clear association with dosage or duration of treatment. Stimulation of hypersensitive dopaminergic receptors in the nigrostriatal system by dopaminergic drugs may explain psychosis early in the disease, but it is unlikely to explain late psychosis. The therapeutic efficacy of atypical neuroleptics suggests a role for mesolimbic dopaminergic and serotonergic pathways. Cholinergic deficiency may also be relevant, in patients with dementia and atrophy of the nucleus basalis.

(c) Management

In patients with clouded consciousness, infection, cerebrovascular accidents, and other relevant pathologies need to be excluded. Revision of dopaminergic medication should come next, with reduction of polypharmacy and dose tapering. Anticholinergics, selegiline, amantadine, and dopamine agonists may need to be discontinued, as they are more likely to trigger psychosis.

(d) Treatment

In most patients antipsychotic drugs are needed, as dopaminergic medication is needed to preserve acceptable motor function. Atypical neuroleptics such as clozapine, with its low D2 receptor affinity and few extrapyramidal side-effects, are preferable to typical neuroleptics. Clozapine is effective at doses of less than 100 mg per day. Initial recommended doses of 6.25 to 12.5 mg daily should be gradually increased until the symptoms are controlled. The danger of agranulocytosis and the need to monitor the blood picture

are the main drawbacks, and other atypical neuroleptics (particularly quietiapine) may be preferable. Cholinesterase inhibitors may also be helpful in treating psychotic symptoms in patients with cognitive impairment.

Dopamine dysregulation syndrome

The syndrome is characterized by the compulsive use of dopaminergic medication beyond that needed to control motor symptoms. It is commoner in males with young onset and may affect 4 per cent of patients. Drug-hoarding and drug-seeking behaviour, impaired social functioning, aggression, and reluctance to reduce medication despite severe dyskinesias are common features. Hypomania and frank psychosis may follow.[14]

(a) Management

Reduction of medication resolves symptoms, but a withdrawal state characterized by dysphoria and irritability follows. Treatment is difficult and often unsuccessful, and primary prevention is preferable.

Tourette syndrome

The syndrome is characterized by motor and phonic tics of fluctuating severity.

- Motor tics appear early, between the ages of 3 and 8 years. Simple tics, e.g. eye blinking, are followed by complex stereotypies (i.e. touching, licking). Phonictics (sniffing, throat clearing) appear later.
- Severity of tics peaks by the age of 20 and may lessen thereafter, but total recovery is rare. Severe cases may start in adulthood.
- Tics, preceded by premonitory urges, occur many times a day and are exacerbated by anxiety, boredom, fatigue, and excitement and lessened by alcohol, relaxation, and sleep. Patients may be able to suppress tics for long periods at the expense of a build-up of tension.
- Coprolalia (utterance of obscenities), copropraxia (obscene gestures), echolalia and echopraxia (repetition of words and actions), and self-harm are also common.

Tourette syndrome is part of a spectrum that includes transient childhood tic disorders. Secondary tic disorders following trauma, encephalitis, rheumatic fever, and metabolic and toxic encephalopathies, and those present in inherited degenerative conditions such as Huntington's disease and neuroacanthocytosis, need to be considered in the differential diagnosis.

Epidemiology

Tourette syndrome is more frequent in males (4:1). Its prevalence in male adolescents is around 4 per 10 000, but it may be higher (49 per 10 000) in children with behavioural disturbances[15] and is as high as 1 to 3 per cent in school children if a broad definition of chronic motor and phonic tics is used.

Aetiology

Major gene effects with an autosomal mode of inheritance seem likely. Monozygotic twins have a concordance rate between 50 and 70 per cent for the syndrome compared with 10 to 20 per cent for dizygotic twins, and a third of relatives may have features of the syndrome. Several candidate genes have been assessed including dopamine receptors, the dopamine transporter, and various noradrenergic and serotonergic genes, and although isolated changes in a single locus are unlikely to lead to the syndrome, these alleles could have a significant cumulative effect. Gestational and perinatal risk factors may also play a role. Multiple streptococcal infections in the months preceding symptom onset are commoner in Tourette patients than in controls, suggesting that the paediatric autoimmune neuropsychiatric disorder associated with streptococcal infection (PANDAS) may be the cause of Tourette syndrome in vulnerable patients.

The efficacy of dopamine antagonists and SSRIs in the treatment of tics and obsessive–compulsive symptoms has implicated dopaminergic and serotonergic pathways and the regions where dopaminergic and serotonergic neurons interact (i.e. striatum, substantia nigra, and prefrontal cortex). The beneficial effects of basal ganglia deep brain stimulation also point to basal ganglia dysfunction. PET studies have suggested decreased metabolism and blood flow in basal ganglia–thalamo-cortical projection systems.[16] Post-mortem studies have not shown consistent abnormalities in D1/D2 receptors, and *in vivo* receptor-binding studies have been conflicting, as have structural brain-imaging studies.

Psychiatric comorbidity

(a) Clinical features

- About half of the patients meet criteria for other psychiatric disorders,[17] but it is uncertain whether they should be considered part of the phenotype.
- Depression and anxiety occur in about 25 per cent of patients.
- Personality disorder is present in two-thirds of patients. Borderline, obsessive–compulsive, and paranoid types are the commonest types.
- Attention-deficit hyperactivity disorder may be more common in males and may have educational and behavioural implications. Wide variations in prevalence have been reported (8 to 80 per cent).
- Obsessive–compulsive disorder may be more common in females, reaching its peak in late adolescence. Concern for symmetry, violent and sexual thoughts, forced touching, fear of harming self and others, and a need to do things 'just right' are common features.
- Intellectual ability tends to be normal, but poor performance in complex attentional tasks is associated with attention-deficit hyperactivity disorder.

(b) Management

Explanation and reassurance are often enough in mild cases; for the rest, drug treatment or cognitive-behaviour therapy may be indicated.

(c) Treatment

Neuroleptics are useful in treating tics. Haloperidol, pimozide, and sulpiride are commonly prescribed and atypical neuroleptics (e.g. risperidone, zipresidone, and olanzapine) are also useful, although few controlled trials are available. Clonidine, a $\alpha 2$ adrenergic agonist, is useful, with less severe side-effects. Behavioural interventions (e.g. habit reversal training and bio-feedback techniques) may also suppress tics. Deep brain stimulation of the

centromedian-parafascicular complex of the thalamus or the internal segment of the globus pallidus is reported to ameliorate tics and self-harm.

Variable success has been achieved with behavioural techniques, SSRIs (fluoxetine), and risperidone in the treatment of obsessive–compulsive disorder. The treatment of attention-deficit hyperactivity disorder is more controversial, but benefits may follow the use of clonidine and stimulants such as methylphenidate, pemoline, and dextroamphetamine without increasing tic severity. (See also Chapter 9.2.4.)

Multiple sclerosis

Multiple sclerosis is a common neurological disease, with a prevalence of 50 to 60 per 100 000; it is more common in women. It usually starts between the ages of 20 to 40 and is characterized by multiple demyelinating lesions with a predilection for the optic nerves, cerebellum, brainstem, and spinal cord. In most cases the disease initially follows a relapsing–remitting course, entering a secondary, progressive phase after some years. For a few patients the disease is progressive from the outset.

Purely psychiatric presentations of multiple sclerosis other than dementia are rare, but psychiatric symptoms are common in the course of the illness.

Depression

- Depressive symptoms occur in about 50 per cent of patients in cross-sectional studies[18] and their lifetime prevalence is also around 50 per cent.
- The rates of suicide are more than twice those of the general population, and young males and those socially isolated or with drinking problems are at a greater risk.
- Low mood, negative thoughts, anhedonia, and suicidal ideation are common features of depression.
- Fatigue and poor concentration may be features of multiple sclerosis and have less diagnostic value.

Euphoria

- It is only present in about 10 per cent of patients and is characterized by mild, continuous elation.
- It is best considered as an organic type of personality change.

Emotional lability

- It is as frequent as euphoria.
- Excessive crying is more frequent than laughter.
- It tends to be more severe in those with significant depression.[19]

Psychosis

- It is uncommon, but brief affective or schizophrenia-like psychoses may occur in patients with well-established multiple sclerosis, sometimes coinciding with a relapse.[20]
- Persecutory delusions and lack of insight are common.
- In most patients these are single episodes lasting 4 to 6 weeks that respond well to symptomatic treatment.

(a) Mechanisms of psychotic symptoms

Severity of brain disease, as measured by magnetic resonance imaging (**MRI**), and duration of illness are not closely correlated with **depression**, but they are a risk factor. The personal and social limitations imposed by the disease are an important risk factor for depression.[18] A **genetic predisposition** has been reported in multiple sclerosis patients with bipolar illness.

Euphoria and **emotional lability** tend to occur in patients with advanced disease and cognitive impairment and are more closely related to MRI indices of brain damage. MRI lesions tend to cluster around the temporal lobes in patients with **psychotic symptoms**.

(b) Management of psychotic symptoms

All patients should be assessed for the presence of depression. Fatigue and poor concentration have limited diagnostic value, as they may be features of multiple sclerosis. Although disease-modifying treatments do not increase the overall risk of depression, they may do so in the first 6 months of treatment in those with a previous history of depression.[21] Regular psychiatric assessment in the early stages of treatment is, therefore, important.

(c) Treatment of psychotic symptoms

Few studies have assessed the effect of antidepressants in patients with multiple sclerosis, but SSRIs and other antidepressants appear to be effective. Their side-effects need to be carefully considered for their potential to aggravate or improve neurological symptoms. Cognitive-behaviour therapy aimed at improving coping strategies is also useful. **Emotional lability** responds well to small doses of SSRIs or tricyclic antidepressants but tends to recur when these drugs are discontinued.

Psychotic episodes may require the use of neuroleptics for brief periods, but the long-term use of these drugs is rarely required.

Cognitive impairment

Cognitive impairment is present in about 40 per cent of multiple sclerosis patients[22] and contributes significantly to the overall disability. The pattern of impairment is characterized by the following:

- Attention deficits and slowing of information processing speed are often the first manifestations and may be present early in the disease.[23]
- Memory disturbances, with greater impairment of recall over recognition, appear later.
- Executive function deficits, with poor working memory, abstract reasoning, and use of strategy, are common.
- Language skills and visuospatial functions tend to be preserved.

Cognitive impairment is greater in those with progressive, severe disease, although cases presented as dementia have also been described.[24] Depression worsens cognitive impairment by slowing down information processing and interfering with learning and working memory. Cognitive impairment correlates with MRI markers of disease severity, in particular brain atrophy.

(a) Management of cognitive impairment

Disease-modifying treatments may slow down cognitive impairment, but evidence is insufficient. The same applies to cholinesterase inhibitors. Cognitive rehabilitation has so far been disappointing.

Space-occupying lesions

Brain tumours

Their clinical manifestations are determined by location and by the effects of raised intracranial pressure. Psychiatric symptoms occur in 50 per cent of patients[25] and are of three main types:

- Confusional states and/or progressive cognitive deterioration occur in a third of patients. Disorientation with clouding of consciousness, euphoria, apathy, and loss of insight are prominent in those with confusional states. Progressive memory impairment, loss of initiative, and bradyphrenia occur in patients with a more protracted course and may coexist with signs of raised intracranial pressure.
- Behavioural and mood disturbances occur in 20 per cent of patients. Irritability, euphoria, depression, and, less frequently, psychosis are part of the picture.
- Paroxysmal disturbances such as poorly formed visual hallucinations and automatisms, indicating temporal lobe involvement, are less common.

Fast-growing tumours are more likely to cause psychiatric symptoms (60 per cent in patients with gliomas and 42 per cent in those with meningiomas). Frontal lobe tumours may present with psychiatric symptoms in the absence of other neurological abnormalities.[26] Medial and orbitofrontal tumours lead to emotional symptoms while disinhibition and irritability or marked apathy occurs when the anterior cingulate is involved. Tumours involving the dorsolateral prefrontal regions are more likely to produce abnormalities of executive function (planning, goal-directed behaviour, ability to monitor effective performance). Disturbances of micturition are specifically associated with frontal tumours.

(a) Management

Brain tumours should be suspected in patients with the above syndromes or when psychiatric symptoms are accompanied by neurological abnormalities or appear *de novo* late in life. Imaging usually confirms the diagnosis.

Neurofibromatosis

There are two types of neurofibromatosis, both inherited as autosomal dominant disorders.

Neurofibromatosis 1 is the commonest (incidence 1/3000) and is characterized by cutaneous manifestations (café-au-lait pigmentation) and neurofibromas, which are benign nerve sheath tumours. Gliomas and hamartomas, especially in the eye and optic pathways, and bone dysplasia are also features. Severe learning disability occurs in 4 per cent of patients, but milder cognitive impairment is commoner (80 per cent of cases). Attentional difficulties are common and a third of patients fulfil criteria for attention-deficit hyperactivity disorder. Perceptual and executive functions are worse than memory functions.[27, 28] Epilepsy is twice as common as in the general population. Ventricular enlargement and T_2 MRI hyperintensities in the basal ganglia, internal capsule, thalamus, brainstem, and cerebellum not closely related to the severity of the learning disability are often present.

Neurofibromatosis 2 is much less common (incidence 1/40 000) and is characterized by bilateral vestibular schwannomas, which may also occur in other peripheral nerves. Meningiomas and ependymomas also occur. Hearing loss, vestibular disturbances, and cataracts are the commonest clinical presentations. Cognition is normal.

The loci for neurofibromatosis 1 and 2 have been located to 17q 11.2 and 22q 12.2 respectively. The rate of spontaneous mutations is high and both genes may have tumour-suppressant roles.

Management is aimed at dealing with attention-deficit hyperactivity disorder, learning disability, and epilepsy.

Tuberous sclerosis

Tuberous sclerosis is a rare autosomal dominant disorder with a prevalence of 1/27 000, variable expressivity, and a high spontaneous mutation rate. There is genetic heterogeneity, with loci described in chromosomes 9 and 16. It is characterized by the presence of skin lesions (adenoma sebaceum), calcified subependymal nodules (tubers), and cortical dysplasias. Hamartomas and other neoplasms of the brain, heart, kidney, and liver are part of the picture.

- **Epilepsy** occurs in 60 to 80 per cent of patients, and infantile spasms, a type of epilepsy, with onset in the first 6 months of life, are particularly common.
- Moderate to severe **learning disability** is present in over 50 per cent of patients and is associated with the presence of infantile spasms and poorly controlled epilepsy.[29]
- **Autism** is 200 times more common in tuberous sclerosis than in the general population and tuberous sclerosis occurs in 1 per cent of autistic patients.[30]
- The number of cortical tubers detected by MRI is a marker of disease severity and is related to the degree of learning disability.[31] The presence of tubers in the temporal lobes has been reported to be associated with autism in patients with tuberous sclerosis.

Management is aimed at control of epilepsy and learning disability.

Further information

Lishman, W.A. (1997). *Organic psychiatry: psychological consequences of cerebral disorder* (3rd edn). Blackwell, London.

Mitchell, A.J. (2003). *Neuropsychiatry and behavioural neurology explained*. WB Saunders, Edinburgh.

Rickards, H. (2005). Depression in neurological disorders: Parkinson's disease, multiple sclerosis, and stroke. *Journal of Neurology, Neurosurgery, and Psychiatry*, **76**(Suppl. 1), 48–52.

Burn, D.J. and Tröster, A.I. (2004). Neuropsychiatric complications of medical and surgical therapies for Parkinson's disease. *Journal of Geriatric Psychiatry and Neurology*, **17**, 172–80.

Siegert, R.J. and Abernethy, D.A. (2005). Depression in multiple sclerosis: a review. *Journal of Neurology, Neurosurgery and Psychiatry*, **76**, 469–75.

Amato, M.P., Zipoli, V., and Portaccio, E. (2006). Multiple sclerosis-related cognitive changes: a review of cross-sectional and longitudinal studies. *Journal of the Neurological Sciences*, **245**, 41–6.

Albin, R.L. and Mink, J.W. (2006). Recent advances in Tourette syndrome research. *Trends in Neurosciences*, **29**, 175–82.

Prather, P. and de Vries, P.J. (2004). Behavioral and cognitive aspects of tuberous sclerosis complex. *Journal of Child Neurology*, **19**, 666–74.

References

1. Chan, K.-L., Campayo, A., Moser, D.J., *et al.* (2006). Aggressive behavior in patients with stroke: association with psychopathology and results of antidepressant treatment on aggression. *Archives of Physical Medicine and Rehabilitation*, **87**, 793–8.

2. Robinson, R.G. (1998). Relationship of depression to lesion location. In *The clinical neuropsychiatry of stroke: cognitive, behavioural and emotional disorders following vascular brain injury*, pp. 94–124. Cambridge University Press.
3. Sharpe, M., Hawton, K., Seagroatt, V., *et al.* (1994). Depressive disorders in long-term survivors of stroke: associations with demographic and social factors, functional status, and brain lesion volume. *The British Journal of Psychiatry*, **164**, 380–6.
4. Narushima, K., Chan, K.-L., Kosier, J.T., *et al.* (2003). The effect of early versus late antidepressant treatment on physical impairment associated with poststroke depression: is there a time-related therapeutic window? *The Journal of Nervous and Mental Disease*, **191**, 645–52.
5. Mayberg, H.S. and Solomon, D.H. (1995). Depression in Parkinson's disease: a biochemical and organic viewpoint. *Advances in Neurology*, **65**, 49–60.
6. House, A., Dennis, M., Mogridge, L., *et al.* (1991). Mood disorders in the year after stroke. *British Journal of Psychiatry*, **158**, 83–92.
7. House, A., Dennis, M., Molyneau, A., *et al.* (1989). Emotionalism after stroke. *British Medical Journal*, **198**, 991–4.
8. Brown, R. and Jahanshahi, M. (1995). Depression and Parkinson's disease: a psychosocial viewpoint. *Advances in Neurology*, **65**, 61–84.
9. Berney, A., Vingerhoets, F., Perrin, A., *et al.* (2002). Effect on mood of subthalamic DBS for Parkinson's disease: a consecutive series of 24 patients. *Neurology*, **59**, 1427–9.
10. Starkstein, S.E., Mayberg, H.S., Leiguarda, R., *et al.* (1992). A prospective longitudinal study of depression in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, **55**, 377–82.
11. Menza, M.A., Robertson–Hoffman, D.E., and Bonapace, A.S. (1993). Parkinson's disease and anxiety: comorbidity with depression. *Biological Psychiatry*, **34**, 465–70.
12. Remy, P., Doder, M., Lees, A., *et al.* (2005). Depression in Parkinson's disease: loss of dopamine and noradrenaline innervation in the limbic system. *Brain: a Journal of Neurology*, **128**, 1314–22.
13. Paulus, W. and Jellinger, K. (1991). The neuropathologic basis of different clinical subgroups of Parkinson's disease. *Journal of Neuropathology and Experimental Neurology*, **50**, 743–55.
14. Giovannoni, G., O'Sullivan, J.D., Turner, K., *et al.* (2000). Hedonistic homeostatic dysregulation in patients with Parkinson's disease on dopamine replacement therapies. *Journal of Neurology, Neurosurgery, and Psychiatry*, **68**, 423–8.
15. Robertson, M.M. and Stern, J.S. (1997). The Gilles de la Tourette syndrome. *Critical Reviews in Neurobiology*, **11**, 1–19.
16. Tanner, C.M. and Goldman, S.M. (1997). Epidemiology of Tourette syndrome. *Neurologic Clinics*, **15**, 395–402.
17. Eidelberg, D., Moeller, J.R., Antonini, A., *et al.* (1997). The metabolic anatomy of Tourette's syndrome. *Neurology*, **48**, 927–34.
18. Ron, M.A. and Logsdail, S.J. (1989). Psychiatric morbidity in multiple sclerosis. A clinical and MRI study. *Psychological Medicine*, **19**, 887–95.
19. Dark, F.L., McGrath, J.J., and Ron, M.A. (1996). Pathological laughing and crying. *The Australian and New Zealand Journal of Psychiatry*, **30**, 472–9.
20. Feinstein, A., du Boulay, G., and Ron, M.A. (1992). Psychotic illness in multiple sclerosis. A clinical and magnetic resonance imaging study. *British Journal of Psychiatry*, **161**, 680–5.
21. Goldman Consensus Group. (2005). The Goldman consensus statement on depression in multiple sclerosis. *Multiple Sclerosis*, **11**, 328–37.
22. Rao, S.M., Leo, G.J., Ellington, L., *et al.* (1991). Cognitive dysfunction in multiple sclerosis. II: impact on employment and social functioning. *Neurology*, **41**, 692–6.
23. Feinstein, A., Youl, B., and Ron, M. (1992). Acute optic neuritis: a cognitive and magnetic resonance imaging study. *Brain: a Journal of Neurology*, **115**, 1403–15.
24. Hotopf, M.H., Pollock, S., and Lishman, W.A. (1994). An unusual presentation of multiple sclerosis. Case report. *Psychological Medicine*, **24**, 525–8.
25. Hecaen, H. and Ajuriaguerra, J. (1956). *Troubles mentaux au cours des tumeurs intracraniennes*. Masson, Paris.
26. Ron, M.A. (1989). Psychiatric manifestations of frontal lobe tumours. *British Journal of Psychiatry*, **155**, 735–8.
27. Hyman, S.L., Shores, A., and North, K.N. (2003). The nature and frequency of cognitive deficits in children with neurofibromatosis type 1. *Neurology*, **65**, 1037–44.
28. Yohay, K. (2006). Neurofibromatosis types 1 and 2. *The Neurologist*, **12**, 86–93.
29. Jozwiak, S., Goodman, M., and Lamm, S.H. (1998). Poor mental development in patients with tuberous sclerosis complex: clinical risk factors. *Archives of Neurology*, **55**, 379–84.
30. Fombonne, E., Du Mazaubrun, C., Cans, C., *et al.* (1997). Autism and associated medical disorders in a French epidemiological survey. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**, 1561–9.
31. Joinson, C., O'Callaghan, F.J., Osborne, J.P., *et al.* (2003). Learning disability and epilepsy in an epidemiological sample of individuals with tuberous sclerosis complex. *Psychological Medicine*, **33**, 335–44.

5.3.3 Epilepsy

Brian Toone

Introduction

An epileptic seizure has been defined as 'a clinical manifestation presumed to result from an abnormal and excessive discharge of a set of neurones in the brain'.[1] A diagnosis of epilepsy applies with the recurrence of two or more discrete and unprovoked seizures (febrile and neonatal seizures are excluded from this definition).

Epilepsy is one of the more common neurological disorders. It carries with it a greater psychiatric morbidity than is to be found in other neurological disorders of comparable severity. Many of its manifestations resemble and may be confused with psychiatric phenomenology. It is often associated with learning difficulties; it may be a manifestation of acquired brain damage or disease; seizures may occur in the course of substance abuse or be caused by psychiatric treatment. For these and for many other reasons psychiatrists should be familiar with epilepsy, its manifold aetiologies, presentations, and treatment.

Classification: epileptic seizures and epilepsy

A comprehensive taxonomy should embrace classifications of both seizure semiology (i.e. the manifestations of abnormal discharge activity) and of epilepsy syndromes. The position of each seizure type in a seizure classification system is determined by its clinical manifestations, by electroencephalographic changes during the seizure, and by the interictal electroencephalographic abnormalities. A classification of epilepsy syndromes takes into account seizure subtype, and also anatomical substrate, aetiology, age of onset, and other characteristics. Seizure classification is dependent upon entities that are immediately ascertainable; epilepsy syndrome classification depends upon entities (e.g. neuroanatomical substrate) that are more speculative. The International League Against Epilepsy

(ILAE) has chosen to give the former priority, while recognizing the importance of the latter.

A familiarity with terminology, with the more common seizure subtypes, and the more commonly encountered epileptic syndromes, will assist in an understanding of the psychiatric disorders that occur in patients with epilepsy. The aura is a simple partial seizure, i.e. a seizure of focal onset in which consciousness is retained. It may progress to a complex partial seizure in which consciousness is disturbed, or into a generalized tonic, clonic, or tonic–clonic seizure. It may subside without further development. It rarely lasts for more than a few seconds, although patients often find time estimation difficult. It is to be distinguished from the epileptic prodrome, a period characterized by dysphoria, impaired memory and concentration, and minor motor manifestations, which precede the seizure and may last for hours or even days. An automatism may be defined as 'a state of clouding of consciousness which occurs during or after a seizure, and during which the individual retains the control of posture and muscle tone and performs simple or complex movements without being aware of what is happening'.[2] The initial phase, consisting of staring or simple chewing movements, may progress to more complex, stereotyped, and repetitive movements such as fumbling or picking. Automatisms rarely last more than a few minutes and are often very brief. They usually arise from temporal lobe discharges, but may be associated with orbital and mesial frontal lesions. The ictus refers to the period of manifest seizure activity. If this persists for 30 min or more it is described as status epilepticus and constitutes a medical emergency (Table 5.3.3.1).

Seizure types

Partial seizures

Partial seizures have a focal onset and may or may not generalize. They may be simple or complex, depending upon whether or not consciousness remains undisturbed. For this purpose, consciousness is defined as the ability to remain aware or to respond.

(a) Simple partial seizures

The content of the simple partial seizure depends upon the site of the focus. One that arises from motor territory may present as a Jacksonian 'march' or as a versive turning of the head. Speech arrest may be present. A focus in the primary sensory cortex may give rise to poorly formed sensations. The more elaborate sensations and the psychic symptoms that arise, respectively, from the association cortex and from the mesial frontotemporal structures are more likely to progress to complex partial seizures.

(b) Complex partial seizures

These may begin as a simple partial seizure or 'aura', or consciousness may be impaired from the beginning. Characteristic auras include epigastric sensations rising into the thorax and olfactory and gustatory hallucinations, elaborated auditory and visual hallucinations, complex changes in perception (e.g. micropsia, depersonalization), and psychic phenomena such as *déjà vu*. As discharge activity spreads, automatic behaviour may supervene or a secondary generalized seizure may ensue. Complex partial seizures commonly arise from temporal, particularly mesial temporal, structures. Hence the obsolete term 'temporal lobe epilepsy'. They may also arise from the orbital and mesial frontal cortices. Complex partial status, previously known as temporal lobe status, is uncommon. It may present as an organic confusional state and may be mistaken for a florid psychosis. Electroencephalography will usually confirm the diagnosis.

Generalized seizures

Both hemispheres are initially and simultaneously involved, the ictal electroencephalographic pattern and motor manifestations are bilateral and consciousness may be impaired from the onset.

Tonic–clonic seizures

This is the common 'major' seizure formerly referred to as 'grand mal'. A brief tonic phase leads into clonic activity, the entire seizure lasting about 2 min. Generalized seizures may be primary, arising from both hemispheres simultaneously, or they may be due to secondary generalization from a focal onset.

Absence seizures

An absence attack is characterized by abrupt cessation of ongoing activity, a vacant stare, and a period of unresponsiveness lasting from a few seconds to half a minute. The absence may be accompanied by brief clonic movements, especially of the eyelids, a reduction in tone causing the body to slump, or automatic movements. Absence seizures occur more commonly during childhood. The absence must be distinguished from the complex partial seizure, which it may resemble. Absence status is not uncommon in childhood and may be mistaken for inattention.

Myoclonic seizures

These are brief shock-like contractions of groups of muscles.

Table 5.3.3.1 Classification of seizures

1 Localization (partial, focal) seizures
(a) Simple partial seizures
(b) Complex partial seizures
(c) Partial seizures evolving to secondary generalized seizures
2 Generalized seizures
(a) Absence seizures
(b) Myoclonic jerks
(c) Clonic seizures
(d) Tonic seizures
(e) Tonic–clonic seizures
(f) Atonic seizures

The epilepsy syndromes

The ILAE, in order to embrace a wider range of clinical features than is possible in a classification based on seizure types, introduced a classification of seizures and epilepsy syndromes. In this classification the broad division lies between the localization-related or partial epilepsies, the great majority of which, in adults, is made up of symptomatic epilepsies (i.e. epilepsy arising from a known or suspected cause) and the generalized epilepsies. The characteristics of the partial epilepsies are determined by the function of the cortical site from which the seizure emanates. The more common generalized epilepsies are age-related. Though admirable in concept, the classification has proved unwieldly and

the terminology clumsy: as such it seems unlikely to displace the seizure-type classification.

Epidemiology

The cumulative incidence (i.e. lifetime risk) is 3.4 per cent for males, 2.8 per cent for females, but the prevalence is only 7 per 1000. This is because prevalence represents the balance between newly diagnosed cases and permanent remission or death, and epilepsy has a good prognosis with 76 per cent of newly diagnosed cases entering long-term remission.[3] Approximately half can be classified as partial seizures and 40 per cent as generalized. Prognosis varies according to the epilepsy subtype, with partial epilepsy having a poorer outcome. Consequently, people who attend a hospital clinic are unrepresentative, in that they are more likely to have treatment-refractory partial seizures along with the other adverse prognostic factors such as mental handicap, neurological dysfunction, or psychiatric disorders.

Aetiology

Aetiology varies according to the age of onset. In childhood- and adolescence-inherited disorders of metabolism, ante- and perinatal complications, infection, migrational errors, and the consequences of febrile convulsions predominate, in middle life trauma and tumour are most common, and in advanced years cerebrovascular disease and degenerative disorders are predominant.

Only one-quarter to one-third of cases of epilepsy are due to known causes. Many others fall into recognizable syndromes about which much is understood. The partial epilepsies are, by definition, due to focal areas of damage and dysfunction usually involving the cortex. However, although the site may be suggested by the seizure semiology, comprehensive investigation may fail to identify any abnormality. Even when it does so, the radiological appearance may lack aetiological specificity. Some generalized seizures may be identified as primary generalized epilepsy syndromes; for example, juvenile myoclonic epilepsy. These are of uncertain aetiology, though genetic factors are considered important; onset is in childhood or adolescence and the prognosis favourable.

In psychiatric practice seizures may arise iatrogenically; they are usually due to pharmacotherapy, less commonly to electroconvulsive therapy. They may result from the overhasty withdrawal of benzodiazepines or to the use of antidepressant or antipsychotic drugs, most of which are epileptogenic. Such seizures are thought to be provoked and do not form grounds for a diagnosis of epilepsy. Adjustment of drug dosage is usually all that is required. Provoked seizures may also occur during alcohol intoxication ('rum fits') or withdrawal; a genetic predisposition may play a part.

Diagnosis

The diagnosis of epilepsy depends first and foremost on historical information; the patient's own account of the seizure and the observations of a reliable informant are of tantamount importance. A family history of epilepsy should be sought; age of onset should be determined when possible. A history of birth complications, febrile fits, early head injury, or cerebral infection is of particular importance in seizures starting in childhood, adolescence, or early adult life. In middle life symptoms suggestive of developing intracranial pathology and in later life cerebrovascular and degenerative disorders should be sought. The clinician should be aware of specific circumstances and situations that may provoke seizures: alcohol or substance abuse, prescribed drugs that have epileptogenic properties, and intermittent photic stimulation. Physical examination will detect not only gross congenital abnormalities such as tuberosclerosis, but also more subtle features, for example facial or skull asymmetries. The differential diagnosis varies according to age group, but will include vasovagal attacks and pseudoseizures, particularly in the young, vertigo and transient ischaemic attacks in the elderly, and cardiogenic syncope, hypoglycaemic episodes, and migraine at any age.

The role of physical investigation is to confirm the diagnosis of epilepsy when this is in doubt and to identify the cause; it may also help to determine the type of epilepsy and, in the partial epilepsies, the site of seizure onset. Magnetic resonance scanning is now widely available and all new cases of adult-onset epilepsy and patients of any age with partial epilepsy should have the benefit of this technology. A sleep electroencephalograph is still mandatory and may be invaluable not only in differential diagnosis, but in determining the epilepsy subtype and seizure localization. Video-telemetry is invaluable when there is continuing diagnostic uncertainty, and in those cases in which precise localization is necessary for presurgical assessment. Functional neuroimaging may also aid localization.

In psychiatric practice epileptic seizures must be distinguished from psychogenic pseudoseizures, panic attacks, and aggressive episodes. The pseudoseizure may take different forms: the patient may fall or slump to the ground and remain still as in a syncopal attack; or there may be jerking or thrashing of limbs resembling a major tonic–clonic seizure. The absence of tongue-biting, incontinence, or significant injury is often said to distinguish the pseudoseizure from the epileptic seizure, but such guides are frequently unreliable. A history of other conversion disorders or illness behaviour may be obtained, but pseudoseizures also occur in individuals with epilepsy. A detailed description of the attack from the patient and from a witness will be of the greatest assistance in diagnosis, but when in serious doubt video-telemetry may offer a definitive answer. Other useful investigations include serum prolactin levels, which rise postictally to reach a peak between 20 and 30 min after a major epileptic seizure. Episodes of panic are sometimes mistaken for epileptic seizures. Extreme anxiety, especially when accompanied by hyperventilation, may lead to a subjective diminution in awareness, altered perception, and other features suggestive of complex partial seizures. The context in which the attack occurs, a description of initial autonomic arousal, and other symptoms of anxiety should lead to the correct diagnosis. Sudden outbursts of aggressive behaviour, particularly when out of character and context, often give rise to suspicions of epilepsy. Aggression, especially directed aggression, is extremely unusual as a feature of the epileptic seizure, though it may occasionally be seen during the phase of postictal confusion or postictal psychosis.

The psychiatric consequences of epilepsy

The prevalence of psychiatric morbidity among persons with epilepsy is greater than in the general population, but the increase in prevalence will vary according to the type of epilepsy, the presence and extent of brain damage, and the presence of cognitive and physical disability. The more reliable studies are drawn from

community samples. Of children between the ages of 5 and 14 years, 29 per cent showed some psychiatric disorder compared with 6.8 per cent of the general population. The figure rose to 58 per cent when brain damage was present.[4] Pond and Bidwell[5] surveyed 14 general practices and reported a prevalence of 29 per cent in adults, increasing to over 50 per cent in patients with temporal lobe epilepsy. Neurotic disability accounted for the great majority of cases. Psychiatric morbidity is over-represented in clinic attenders and in patients with partial epilepsy.[6]

Personality disorder and social development

Notions of an epileptic personality arising out of a hereditary 'taint' persisted well into the present century. The person with epilepsy was said to be explosively aggressive, rigid, egocentric, and irritable. These beliefs were formed by observations of often oversedated inmates of epileptic institutions. The concept of a specific epileptic personality has now largely been abandoned, though it is acknowledged that some features associated with, but not specific to, epilepsy may exercise a powerful influence on personality development. Many of these are consequences of brain damage rather than epilepsy as such. Thus learning difficulties, leading to limited educational opportunity, adult unemployment, and socioeconomic disadvantage may be significant personality determinants. But even in the epileptic individual without brain damage the sedative actions of anticonvulsant medication, the continuing stigma of seizure activity, and the social and occupational constraints are not without their effects on the developing personality.

A particular link between temporal epilepsy and abnormalities of personality has long been debated[7] and certain exaggerated traits (e.g. hypergraphia) are consistently reported,[8] but many of the personality difficulties may be explained by the refractory nature of temporal lobe seizures and the need for increased medication.

Psychoses

(a) Chronic interictal psychoses

Throughout the first half of this century the relationship between epilepsy and schizophrenia was debated at length, usually in terms of whether the presence of one condition encouraged or discouraged the development of the other—the affinity and antagonism hypotheses, respectively. In recent years, particularly following the publication of Slater's seminal studies,[9] informed opinion has moved firmly behind the first view. Epidemiological studies based on national registers[10,11] find a higher prevalence of chronic psychosis in epileptic subjects than in the general population. A neurology outpatient clinic study[12] reported schizophrenia to be nine times more common in epilepsy than in a migraine control group. The onset, cause, and clinical characteristics are, to a very large extent, indistinguishable from those of more usual forms of schizophrenia, although negative symptoms occur less frequently, thought disorder is rarely encountered and the outcome may be more benign. Psychosis usually develops 11–15 years following the onset of epilepsy. The aetiology remains uncertain. Cases in which the epilepsy takes the form of complex partial seizures arising from the mesial temporal or frontal lobes are over-represented; there may be a slight left-sided predominance. A family history of schizophrenia was thought to be unusual, but some recent studies have cast doubt on this assumption. Neuropathological examination, more readily available with the increasing practice of epilepsy surgery, has proved less informative than had been hoped, but subjects undergoing temporal lobectomy for resection of small neurodevelopmental lesions, e.g. ganglioglioma, dysembryoplastic neuroepithelioma (DNET)[13] appear at greater risk of developing a post-operative chronic interictal psychosis. The results of structural neuroimaging studies have been inconclusive with reports of both reduction and increase in amygdala size. The risk of bipolar illness or affective psychosis does not appear to be increased in epilepsy.

(b) Postictal psychosis

The other common form of epileptic psychosis develops following an exacerbation of seizure activity. Because of this close temporal relationship the instrumental role of epilepsy in the aetiology of the psychosis is not open to question. The salient characteristics have been described.[14] The psychosis usually occurs following a cluster of complex partial seizures usually followed by secondary generalization. Characteristically, the subject appears to make a complete recovery, but 1 to 2 days later, the so-called lucid interval, becomes floridly psychotic. Affective, schizophrenic, and confusional elements may be present. An electroencephalogram recorded at the time shows increased focal discharge activity, though less than would be seen in partial status. Spontaneous recovery is to be expected, usually within a week of onset. The first episode is usually delayed until early adult life, a decade and a half after the first seizure. Half of those affected will have further similar episodes. Fifteen to 20 per cent will progress to develop a chronic interictal psychosis. Bilateral EEG discharges are seen more commonly in those patients who develop postictal psychosis. Functional neuroimaging at the time of the psychosis demonstrates relative mesial temporal hyperperfusion suggestive of an active process, either continuous seizure discharge or seizure inhibition.

Rarely, the introduction of certain anticonvulsant drugs may seem to precipitate psychotic episodes. This may be associated with rapid seizure control and give rise to the term 'forced normalization'.[15] In recent years vigabatrin has seemed to be the drug most responsible.

Sexual function

A diminution in sexual interest, a decrease in activity, and impaired performance are the most common aspects of sexual dysfunction in epilepsy. Men have been studied more thoroughly than women. In patients receiving antiepilepsy drugs libido may be diminished and erectile potency impaired. Levels of free testosterone, the biologically active hormone, may be diminished. Sperm concentrations may be reduced, morphological abnormalities may occur more commonly and mobility may be reduced. Menstrual irregularities are increased in women with epilepsy and are related to seizure frequency, polytherapy with antiepilepsy drugs, and the use of sodium valproate. Infertility, ovulation, and the polycystic ovarian syndrome occur more commonly. Hyposexuality may be more pronounced in patients with partial epilepsy, but this may simply reflect the refractory nature of partial epilepsy and the greater amount of drugs prescribed.[16,17]

Epilepsy and crime

There is an association between epilepsy and criminal activity. Male epileptics are three times more likely to receive a criminal conviction,[18] in England and Wales between 0.7 and 0.8 per cent of the

prison population suffers from epilepsy, a figure considerably higher than in the general population,[19] but the pattern of offence does not differ. The reasons for this are unclear. Low intelligence and low socioeconomic class are common to both epilepsy and prison populations; the role of brain damage as distinct from epilepsy has not been fully evaluated. Crimes of violence in the context of disturbed ictal or postictal behaviour do occur, but are extremely rare.[20]

Neurotic illness

Neurotic illness, more especially anxiety and depression, largely account for the increased psychiatric morbidity that is to be found in patients with epilepsy. These disorders have few distinctive characteristics. They may be explained by the adverse social, educational, and economic disadvantages that confront people with epilepsy. A phobic anxiety akin to agoraphobia may be seen in some individuals who fear the onset of a seizure in public places. Obsessive–compulsive disorder does not seem to occur more commonly in epilepsy.

Epilepsy and suicide

Suicide is increased fivefold among patients with epilepsy, but is considerably higher among those with temporal lobe epilepsy.[21] Among patients presenting with self-harm, epileptic subjects are over-represented from five- to sevenfold.[22]

Treatment

Seizure control is most effectively achieved through the appropriate use of anticonvulsant drugs. The use of behavioural techniques to inhibit seizure activity holds promise, but is still in its infancy. Surgical treatment is increasingly available, but should only be considered for those patients who have failed repeatedly to respond to drug therapy and who have resectable lesions. Drug treatment should aim to achieve seizure control through the use of a single anticonvulsant drug, thus minimizing unwanted side-effects. This should be possible in the great majority of patients. If a first-choice drug fails, a second-choice drug should be substituted. Polytherapy may be necessary, notably in the management of partial seizures, but should be avoided wherever possible. Most first-line drugs, except phenytoin, are described in Chapter 6.2.6. Phenytoin, although relatively toxic, especially to the cerebellum, is an effective anticonvulsant and still widely prescribed. Serum monitoring is particularly important. More recently introduced drugs include topiramate and levetiracetam. The benzodiazepines, clobazam and clonazepam, may be used in adjunctive therapy. Lamotrigine and levetiracetam appear to be as efficacious as the longer established drugs and have less side-effects. There is little place for either phenobarbitone, which is unduly sedative and may cause depression and behavioural disturbance especially in children and adolescents, or vigabatrin, which may cause visual field constrictions. Vigabatrin, and to a lesser extent topiramate may cause psychotic episodes. The need for continuing anticonvulsant therapy should be reviewed by a specialist neurologist once the patient has been free of seizures for 2 years. For a more detailed account of the management of epilepsy and of status epilepticus the reader is referred to Shorvon *et al.*[23]

The treatment and outcome for psychogenic seizures have received increasing attention. The importance of early diagnosis is emphasized.[24] Psychological treatment, particularly cognitive–behavioural therapy, can be effective. Reduction or cessation of symptoms can be achieved in at least half of the cases within a 6–12 month period.

Further information

Asbury, A.K., McKhann, G.M., *et al.* (eds.) (1992). *Diseases of the nervous system* (2nd edn). W.B. Saunders, Philadelphia, PA.

Hopkins, A., Shorvon, S., and Cascino, G. (eds.) (1995). *Epilepsy* (2nd edn). Chapman & Hall Medical, London.

Lishman, W.A. (1998). *Organic psychiatry* (3rd edn). Blackwell Science, Oxford.

Trimble, M.R. (1991). *The psychoses of epilepsy*. Raven Press, New York.

References

1. Hopkins, A. and Shorvon, S. (1995). Definitions and epidemiology of epilepsy. In *Epilepsy* (2nd edn) (eds. A. Hopkins, S. Shorvon, and G. Cascino), pp. 1–25. Chapman & Hall Medical, London.
2. Fenton, G.W. (1981). Psychiatric disorder of epilepsy: classification and phenomenology. In *Epilepsy and psychiatry* (eds. E.H. Reynolds and M.R. Trimble), pp. 12–26. Churchill Livingstone, Edinburgh.
3. Annegers, J.F., Hauser, W.A., and Elveback, L.R. (1979). Remission of seizures and relapse in patients with epilepsy. *Epilepsia*, **20**, 729–37.
4. Graham, P. and Rutter, M. (1968). Organic brain dysfunction and child psychiatric disorder. *British Medical Journal*, **3**, 695–700.
5. Pond, D.A. and Bidwell, B.H. (1960). A survey of epilepsy in fourteen general practices. II. Social and psychological aspects. *Epilepsia*, **1**, 285–99.
6. Edeh, J., Toone, B.K., and Corney, R.H. (1990). Epilepsy, psychiatric morbidity, and social dysfunction in general practice. *Neuropsychiatry, Neuropsychology and Behavioural Neurology*, **3**, 180–92.
7. Bear, D.M. and Fedio, P. (1977). Quantitative analysis of inter-ictal behaviour in temporal lobe epilepsy. *Archives of Neurology*, **34**, 454–67.
8. Waxman, S.G. and Geschwind, N. (1974). Hypergraphia in temporal lobe epilepsy. *Neurology*, **24**, 629–36.
9. Slater, E., Beard, A.W., and Glithero, E. (1963). The schizophrenia-like psychoses of epilepsy. *British Journal of Psychiatry*, **109**, 95–150.
10. Bredkjaer, S.R., Mortensen, P.B., and Parnas, J. (1998). Epilepsy and non-organic non-affective psychosis. National epidemiological study. *British Journal of Psychiatry*, **172**, 235–9.
11. Stefansson, S.B., Olafsson, E., and Hauser, W.A. (1998). Psychiatric morbidity in epilepsy: a case control study of adults receiving disability benefit. *Journal of Neurology, Neurosurgery and Psychiatry*, **64**, 238–41.
12. Mendez, M.F., Grau, R., Doss, R.C., *et al.* (1993). Schizophrenia in epilepsy: seizure and psychosis variables. *Neurology*, **43**, 1073–7.
13. Andermann, L.F., Savard, G., Meencke, H.J., *et al.* (1999). Psychosis after resection of ganglioglioma or DNET: evidence for an association. *Epilepsia*, **40**, 83–7.
14. Logsdail, S.J. and Toone, B.K. (1988). Post-ictal psychoses. *British Journal of Psychiatry*, **152**, 246–52.
15. Landolt, H. (1953). Some clinical EEG correlations in epileptic psychoses (twilight states). *EEG and Clinical Neurophysiology*, **5**, 121.
16. Toone, B.K. (1995). Epilepsy and sexual life. In *Epilepsy* (2nd edn) (eds. A. Hopkins, S. Shorvon, and G. Cascino), pp. 555–65. Chapman & Hall Medical, London.
17. Hertzog, A.G. and Fowler, K.M. (2005). Sexual hormones and epilepsy: threat and opportunities. *Current Opinion in Neurology*, **18**, 167–172.
18. Gudmundsson, G. (1966). Epilepsy in Iceland. A clinical and epidemiological investigation. *Acta Neurologica Scandinavica*, **25**, 7–124.
19. Gunn, J. and Fenton, G.W. (1971). Epilepsy, automatism and crime. *Lancet*, **i**, 1173–6.

20. Delgado-Escueta, A.V., Mattson, R.H., King, L., *et al.* (1981). The nature of aggression during epileptic seizures. *New England Journal of Medicine*, **305**, 7110–16.
21. Barraclough, B. (1981). Suicide and epilepsy. In *Epilepsy and psychiatry* (eds. E.H. Reynolds and M.R. Trimble), pp. 72–6. Churchill Livingstone, Edinburgh.
22. Hawton, K., Fagg, J., and Marsack, P. (1980). Association between epilepsy and attempted suicide. *Journal of Neurology, Neurosurgery and Psychiatry*, **43**, 168–70.
23. Shorvon, S., Dreifuss, F., Fish, D., *et al.* (eds.) (1996). *The treatment of epilepsy*. Blackwell Science, Oxford.
24. Reuber, M., Fernandéz, G. Bauer, J., *et al.* (2002). Diagnostic delay in psychogenic non-epileptic seizures. *Neurology*, **58**, 493–5.

5.3.4 Medical conditions associated with psychiatric disorder

James R. Rundell

Introduction

Seven out of 10 office visits to a primary care practitioner are related to a chronic illness.[1] There are high levels of association of many of these chronic conditions with psychiatric disorders.[1] Comorbid medical and psychiatric conditions increase use of medical resources and costs, as well as amplify functional impairment.[2] For example, depression is associated with an approximately 50 per cent increase in medical costs of chronic medical illness, even after controlling for severity of physical illness.[2, 3] Dementia is associated with hospital costs up to 75 per cent higher than for non-demented patients.[4]

As important as a comprehensive knowledge of psychiatric diagnosis and psychosocial formulation is to a consulting psychiatrist, it is also vital to understand the pathophysiology and clinical characteristics of the medical and surgical conditions that frequently coexist with psychiatric disorders. It is also important to know the behavioural and psychiatric side effects of medications and substances. Lacking this data permits only a partial and inadequate approach to diagnosis and treatment.

This section describes general medical disorders associated with psychiatric syndromes. The pathophysiology and clinical characteristics of the medical disorder are described first, followed by psychiatric syndromes often seen with that diagnosis.

Cardiovascular disorders

Ventricular dysrhythmias

Sudden cardiac death is responsible for 300 000 deaths annually in the United States.[5,6] Sympathetic nervous system activity increases the likelihood of ventricular dysrhythmias[6] especially when there is prior ischaemic damage. Sympathetic nervous system stimulation, which increases heart rate, can trigger ectopic sites in the myocardium, which override normal conductive pathways, producing potentially fatal dysrhythmias. Either the peripheral sympathetic nervous system or the central nervous system can generate stimuli leading to this phenomenon. Therefore, anxiety and stress may increase the risk of dysrhythmia.[7] Among individuals with pre-existing heart disease or dysrhythmias, activities which precipitate adrenergic discharge may produce ventricular dysrhythmias—for example, public speaking, road rage, and recall of emotionally charged events.[8] In one series of patients, psychological stressors were more reliable triggers of dysrhythmias than physical manoeuvres such as carotid sinus massage, hyperventilation, and the Valsalva manoeuvre.[8] Simple and inexpensive non-pharmacological techniques such as relaxation training, hypnosis, and medication have been shown to improve ventricular dysrhythmias.[8, 9]

Depression has also been associated with lower threshold for ventricular dysrhythmias.[10] Patients with depression exhibit dysregulation of the sympathoadrenal system—hypothalamic corticotropin releasing factor-containing neurons appear to stimulate several autonomic centres involved in regulating sympathetic activity.[11] Smith *et al.*[8] found that deaths within 18 months of a myocardial infarction were concentrated among depressed patients with 10 or more premature ventricular contractions per hour. In this group of patients, 83 per cent of mortality was due to 'arrhythmic deaths'.

Hypertension

More than 60 million Americans have hypertension. The prevalence among whites is about 15 per cent, but is over 25 per cent in the African–American population. 95 per cent of people with hypertension have primary, or idiopathic, hypertension. The remaining 5 per cent have secondary hypertension, due to conditions or substances such as renal disease, steroids, or oral contraceptives. Managing hypertension reduces the morbidity and mortality of the condition. Constitutional and stress-related factors contribute to hypertension. Patients with hypertension, in general, have a more prolonged vasoconstrictive response to psychological stress than patients with normotension,[12] which may result in both short-term and long-term blood pressure elevation due to this interplay of environmental and constitutional factors.[13]

Myocardial infarction

Myocardial ischaemia often leads to myocardial infarction. Acute myocardial infarction can develop at rest or with normal activity. Deaths associated with acute myocardial infarction occur during the first few hours after the onset of symptoms, and are the result of ventricular fibrillation. It is important that patients know the warning signs and seek care promptly when symptoms develop. Unfortunately, as many as 20 per cent of myocardial infarctions are unrecognized. Denial of acute myocardial infarction symptoms and warning signs by individuals, particularly men, are a frequent source of mortality and morbidity. The roles of gender-specific differences in terms of establishing predictors for clinical outcomes is understudied.[14]

The most common precipitant of myocardial ischaemia among patients with pre-existing coronary artery disease is stress.[15] Stress-induced ischaemia is more common than ischaemia induced by physical stressors. Recovery from a myocardial infarction is also highly dependent on psychosocial factors. Ruberman *et al.*[16] demonstrated that postmyocardial infarction patients, who are socially isolated and have high stress levels, have at least four times

the risk of death, compared to their counterparts who have lower levels of stress and isolation. Particularly, lack of a close confidant predicts negative outcomes after myocardial infarction, including further cardiac events.[17] In addition, emotional distress after myocardial infarction is associated with poorer outcomes in terms of quality of life and psychological adjustment.[18]

Depression may occur in 31 per cent of patients admitted for acute myocardial infarction.[19] Presence of major depressive disorder in a patient with cardiac disease has a significant association with morbidity and mortality. Carney *et al.*[20] found that major depressive disorder was the best single predictor of myocardial infarction, angioplasty, and death during the 12 months following cardiac catheterization. Patients with a history of myocardial infarction and major depressive disorder are up to three to five times more likely to die within 6 months of discharge than non-depressed patients following infarction.[19] As to how depression increases risk include hypothalamic–pituitary–adrenocortical and/or sympathoadrenal hyperactivity, diminished heart rate responsivity, ventricular instability, myocardial ischaemia due to stress, and alterations in platelet receptors and reactivity.[21, 22] The data are limited, antidepressant treatment, stress management, and relaxation training in patients with coronary artery disease or myocardial infarction and major depression probably reduces mortality.[23]

Type A and type D personality

Assessment of the patient's personality and behavioural style is important because type A behavioural patterns increase the risk of a myocardial infarction.[24] The type A behaviour pattern includes ambitiousness, aggressiveness, competitiveness, impatience, muscle tenseness, alertness, rapid and emphatic vocal style, irritation, cynicism, hostility, and an increased potential for anger. Very frequently, such individuals are also hard-working 'workaholics' who deny physical or emotional vulnerability. Their self-esteem is often dependent on constant achievement. Unstable cardiac function poses an immediate and ongoing threat to them, and challenges their need to be in control of their environment and bodies. Many clinicians believe that modifying a type A behaviour pattern is an integral part of preventing future myocardial infarctions.[24] Group or individual psychotherapy that reduces type A behaviour and other behavioural risks has been shown to lower the incidence of recurrent infarction and cardiac death in patients with a previous myocardial infarction.[25]

There has been recent attention to the type D personality construct.[26, 27] D behaviour is characterized by inhibition of negative emotions and avoiding social contacts with others. D personality patients may be at increased risk for cardiovascular morbidity and mortality.[26] Cortisol may be a mediating factor for this increased risk. D personality predicts cardiac events after controlling for concurrent stress and anxiety.[27] Studies are needed to validate this personality construct, further define associations with cardiac outcome, and develop treatment approaches for patients with this personality style.

Respiratory disorders

Asthma

Asthma affects between 3 and 5 per cent of the population of the United States. The three hallmarks of the disease are airway inflammation, airway hyperresponsiveness, and a partially reversible airway obstruction. It is one of the classic 'psychosomatic diseases'. Emotional arousal causes changes in airway tone. The severity of an asthma attack is highly correlated with presence of major depressive disorder, panic attacks, general anxiety, and level of fear among children, adolescents, and adults.[28] Asthma patients with psychiatric disorders have worse asthma control, more frequent exacerbations, and worse quality of life than asthma patients without psychiatric disorders.[29] Education, relaxation, biofeedback, and family therapy have each shown efficacy in the management of asthma.[30] Important in the management of asthma is education about the adverse effects of antiasthma medications, which include jitteriness, palpitations, and insomnia. These side effects may require treatment with behavioural and/or psychopharmacological therapies.

Chronic obstructive pulmonary disease

Patients with chronic obstructive pulmonary disease (COPD) have slowly progressive airway obstruction. The course of the disease is punctuated by exacerbations due to pulmonary infection, heart failure, and poor compliance with prescribed therapy. Generally affects middle-aged and older patients. They present with dyspnoea, exercise intolerance, cough, and sputum production. Physical examination reveals lung overinflation, prominent use of accessory muscles to augment respiration, diminished breath sounds, and diffuse wheezing. As with asthma, pharmacological treatments for COPD can cause psychiatric symptoms, especially higher doses of steroid medications. Patients with COPD must stop smoking; pulmonary function declines faster in smokers who develop COPD than non-smokers who develop COPD.

The chronic hypoxia caused by COPD compromises cognition and mood, which, in turn, can produce delirium, mood lability, mood disorders, and restriction in daily activities. Depression is present in 20–60 per cent of COPD patients.[31] Depression adversely affects treatment adherence and may increase risk for poor outcomes. There is considerable evidence that supplemental oxygen improves cognitive function and quality of life.[30] Unfortunately, mood improvement with supplemental oxygen has not been conclusively demonstrated.

Panic attacks are reported in up to 38 per cent of patients with COPD.[32] Benzodiazepines, which are highly effective for controlling panic attacks, have limited usefulness in patients with COPD because they can suppress respiratory function and if used chronically result in tolerance and dependence. Carbon dioxide likely plays a role in promoting panic attacks; carbon dioxide levels increase with COPD disease progression. Antidepressants are useful in patients with COPD who develop panic attacks. Low-dose neuroleptic medications (e.g. 0.5–1.0 mg risperidone orally two to three times daily) are also sometimes used for severe fear and panic, especially in intensive care unit settings (e.g. when weaning the patient from a respirator). Neuroleptics do not directly suppress respiration, though caution must be exercised so that the sedation induced by neuroleptics—potentially combined with other sedating agents—does not reduce respiratory effort beyond that required to maintain adequate oxygenation. Function must also be monitored to ensure that neuroleptic use does not affect cardiac conduction or cause dysrhythmias.

Pulmonary embolism

Patients with psychiatric disorders, including bipolar disorder, anxiety disorder, and schizophrenia, are at increased risk for pulmonary embolism.[33] Embolism may account for a portion of the

excess risk of death among people with schizophrenia, even after controlling for blood pressure, cholesterol, body mass index, smoking, exercise, alcohol intake, and education level.[34] Most thromboemboli originate in the deep veins of the thigh. The diagnosis of pulmonary embolus is often missed because the clinical findings are non-specific. They include dyspnoea, pleuritic chest pain, haemoptysis, tachypnoea, and wheezing or crackles on pulmonary examination. Number of factors predispose to pulmonary thromboemboli: cancer, stroke, myocardial infarction, congestive heart failure, sepsis, pregnancy, lower extremity fractures, major surgical procedures, polycythaemia vera, and paroxysmal nocturnal haemoglobinuria. Pulmonary emboli are treated with heparin and warfarin. Fibrinolytic drugs and acute embolectomy are used in certain situations. The differential diagnosis of sudden anxiety or a panic attack includes pulmonary embolus.

Sleep apnoea

Apnoea is defined as the complete cessation of respiratory airflow for 10 or more seconds.[35] Apnoea can occur during any sleep stage, but is particularly likely to occur during the period of rapid eye movement sleep. It is important to remember that normal people have apnoeic episodes during sleep. When apnoeic events are frequent and prolonged, they lead to chronically disrupted sleep and excessive daytime somnolence. This defines the condition known as sleep apnoea. Sleep apnoea can be central, obstructive, or a mixture of the two. Central sleep apnoea is caused by an abnormal central drive to the respiratory muscles. Congestive heart failure is the most common cause, followed by neurological disorders involving the brainstem and respiratory centres. Obstructive sleep apnoea is more common; obesity is a major risk factor, but is not always present. Aside from disrupted sleep and daytime somnolence, associated symptoms include an inability to concentrate, depressed mood, irritability, and personality changes. The sleeping partner often sleeps in another room because of the individual's very loud snoring, snorting, gasping, and restlessness. Treatment with continuous positive airway pressure is often effective. Patients should avoid sedatives and alcohol. If obese, they should lose weight.

Gastrointestinal disorders

Oesophageal dysmotility

Oesophageal dysmotility can be demonstrated in 30 per cent of patients with non-cardiac chest pain;[36] a significant number of non-cardiac chest pain patients lack any evidence of oesophageal reflux and have reduced perception thresholds for pain. Cases of oesophageal dysmotility often lead to psychiatric consultation. Situational stress has not been conclusively linked to oesophageal dysmotility, but major psychiatric illness has.[37] The majority of patients with oesophageal motility disorders have an Axis I psychiatric illness, especially major depressive disorder (52 per cent), generalized anxiety disorder (36 per cent), somatization disorder (20 per cent), and substance-related disorders (20 per cent).[38] Smooth muscle relaxants, such as calcium-channel blockers, are superior to psychiatric treatments in improving physiological measures (such as oesophageal motility testing), antidepressants, and behavioural therapies produce more impressive changes in patients' subjective oesophageal complaints and level of psychological well-being.[39]

Irritable bowel syndrome

Irritable bowel syndrome (IBS) ranks second only to the common cold as a cause of absenteeism from work,[40, 41] affecting between 8 and 17 per cent of the general population in the United States.[40] Symptoms include abdominal pain (relieved by defecation), and various forms of disturbed defecation such as altered stool frequency, altered stool form, altered stool passage, passage of mucus, and bloating. Symptoms must be continuous or recur within 3 months to meet the criteria for a diagnosis of irritable bowel syndrome. The severity of this syndrome frequently correlates with periods of emotional stress; the sympathetic nervous system inhibits gastric motility.

The enteric nervous system contains approximately 100 million neurons, close to the same number found in the spinal cord,[41] and more than those distributed to any other organ or physiological system. It therefore, makes sense that the gastrointestinal tract is uniquely sensitive to the neurophysiological aspects of the stress response. With IBS who seek medical care exhibit high rates of psychiatric disorders. The most frequently occurring are panic disorder (26 per cent), generalized anxiety disorder (26 per cent), social phobia (26 per cent), and major depressive disorder (23 per cent).[42] Patients with irritable bowel syndrome who are depressed and complain of diarrhoea may benefit from tricyclic antidepressant treatment, at least partially because of their anticholinergic effects. Anxious patients may also benefit from, and well-tolerate buspirone. At least one in eight IBS patients are offered an antidepressant,[43] though data suggest that antidepressants are more consistent in improving global measures than specific gastrointestinal symptoms. A group of patients with treatment-refractory irritable bowel syndrome—nearly half had no psychiatric disorder—more than 90 per cent benefited from low-dose antidepressant or antianxiety medications: 92 per cent of patients improved, and 56 per cent experienced complete remission of irritable bowel symptoms.

Inflammatory bowel disease

Inflammatory bowel disease is the collective term for patients who have ulcerative colitis or Crohn's disease. The aetiology of inflammatory bowel disease is unknown, but it may involve immunological, infectious, or environmental factors.[44] The primary manifestations of acute ulcerative colitis are rectal bleeding, diarrhoea, urgency, fever, weight loss, and, sometimes, abdominal pain. Crohn's disease presents with malaise, fever, abdominal pain, and frequently rectal bleeding. Surgical treatment (colectomy) cures ulcerative colitis but not Crohn's disease. However, surgery is usually a last resort in ulcerative colitis.

Despite the strong beliefs of early psychosomatic theorists, there is no objective evidence that psychiatric disorders cause inflammatory bowel disease. However, patients with this disease and who have psychiatric disorders are more likely to have unexplained physical symptoms in other organ systems, more disability than patients with similar disease severity and no psychiatric disorder, and prior histories of physical and sexual abuse.[45] Exacerbations of inflammatory bowel disease symptoms are positively associated with major life events and major stressors.[41,46] Stress-induced alterations in gastrointestinal inflammation may be mediated through changes in hypothalamic–pituitary–adrenal axis function and alterations in bacterial-mucosal interactions, and via mucosal mast cells and mediators such as corticotrophin releasing factor.[47] Treatment focuses on the identification and treatment of

psychiatric disorders, if found, and on stress management and quality of life issues. Walker *et al.*[44] treated inflammatory bowel disease patients who had major depression with an antidepressant and found marked improvement in depression and ability to function. Relaxation, stress management,[45, 48] and hypnotherapy were found to reduce abdominal pain and diarrhoea.

Gastroesophageal reflux and peptic ulcer disease

Acid reflux and peptic ulcer disease are common causes of non-cardiac chest pain.[49] Ulcer disease occurs when the balance between stomach acid and mucosal defence factors is disrupted. Gastric acid, *Helicobacter pylori*, and non-steroidal antiinflammatory drugs are the most important risk factors in the development of peptic ulcers.[50] The majority of patients with peptic ulcer disease are present with epigastric pain that begins 1 to 3 h after eating. Treatment is aimed at reducing gastric acid (e.g. using cimetidine and ranitidine), improving mucosal defences (with, for instance, sucralfate), and/or eradicating *H. pylori* (antibiotics).

On examination, at least half of the patients initially suspected of having peptic ulcer disease do not have evidence of an ulcer.[51] Among patients with non-ulcer dyspepsia, psychiatric comorbidity is high. Magni reported that 87 per cent of patients with non-ulcer dyspepsia have one or more anxiety disorders compared with 25 per cent of those with dyspepsia where there is endoscopic evidence of ulcer.[52] Ang *et al.* reported that a majority of patients who have typical symptoms of gastroesophageal reflux do not have erosions on examination; those patients with non-erosive reflux disease have a higher prevalence of psychiatric morbidity.[53]

Metabolic disorders

Obesity

Obesity is becoming an epidemic throughout the developed world. Existing standard treatments in university settings, only 20 per cent of obese patients lose around 9 kg (about 20 lbs) at 2-year follow-up and only 5 per cent of patients lose about 18 kg (40 lbs).[54] The majority of people who lose weight on a diet gain it all back. Weight loss and weight maintenance after loss is associated with more initial weight loss, reaching a self-determined goal weight, having a physically active lifestyle, a regular meal rhythm including breakfast and healthier eating, control of overeating, and self-monitoring of behaviours.[55] Associated with weight regain after significant weight loss include a history of weight cycling, disinhibited eating, binge eating, more hunger, eating in response to negative emotions and stress, and more passive reactions to problems.

There is no ideal treatment for weight loss. Weight-loss programmes vary considerably in terms of risk, cost, and efficacy. For most patients with mild to moderate obesity, a multidimensional approach is best, combining diet, exercise, behaviour modification, and social support. Motivated patients with morbid obesity (more than 100 per cent overdesired body weight) may be considered for very low calorie diets, with the emphasis on long-term diet, behavioural change, exercise, and social support. Increasingly, surgical approaches are being used; patient selection procedures have been developed to address motivations, psychological resilience, dietary education, potentially complicating psychiatric or substance-related factors, and ensuring patients are aware these procedures are not without risk—there are many structural and metabolic complications, including death. It is important to treat comorbid psychiatric illnesses. It is associated with excessive intake of carbohydrate-rich foods and with resistance to engaging in physical activity.[56] With mood disorders and schizophrenia have a high prevalence of risk factors for cardiovascular disease, diabetes, and obesity, which are on the order of 1.5 to 2.5 times higher than in the general population.[56] On the other hand, it is also important to be mindful of metabolic effects of psychopharmacological agents, especially second generation antidepressants. The latter are associated with weight gain, dyslipidemia, and abnormal glucose homeostasis, especially with olanzapine.[56]

Wilson's disease

Wilson's disease, or hepatolenticular degeneration, is an autosomal recessive disorder affecting between one and three persons per 100 000 of the population. The abnormality in Wilson's disease is defective hepatic excretion of copper. The consequence is copper deposition and injury to many organs, particularly the liver and the brain, including diffuse white matter lesions seen on MRI in many patients.[57] The genetic defect occurs on chromosome 13, and the gene product is probably a transmembrane copper transporter.[58] Because copper accumulation is slow, signs and symptoms do not appear before the age of 6 years. Most patients present with manifestations of organ damage between the ages of 8 and 20. Prolonged extrahepatic release of copper not bound to ceruloplasmin causes basal ganglion destruction, and sometimes cerebral cortex destruction. Prominent neuropsychiatric symptoms include irritability, aggression, disinhibition, and recklessness. Depressive features are also common. The severity of psychiatric symptoms correlates with the severity of neurological symptoms, especially dystonic and bulbar manifestations.[58]

Disorders of lipid metabolism

Intervention studies have shown that cholesterol reduction using diet, drugs, or surgery reduces the risk of developing or worsening coronary disease. In general, a 1 per cent reduction in low-density lipoprotein-cholesterol has been associated with roughly a 2 per cent reduction in disease end-points.[59] General agreement exists that eating less saturated fat and cholesterol, and adopting a diet and exercise habits to reduce obesity will benefit the health of most people. Exercise has a much greater effect in reducing triglyceride levels than in reducing low-density lipoprotein-cholesterol concentrations. Triglyceride levels are reduced after even a single exercise session. The efficacy of regular aerobic exercise in mild to moderate hypertriglyceridaemia has been repeatedly demonstrated.[60]

Hepatic encephalopathy

The pathogenesis of hepatic encephalopathy is related to widespread hepatic necrosis, commonly due to an acute viral infection, such as hepatitis B, or exposure to hepatotoxins. Common hepatotoxins that lead to liver failure include acetaminophen, isoniazid, halothane, valproic acid, mushroom toxin, and carbon tetrachloride. Hepatic encephalopathy that accompanies acute fulminant liver failure is frequently associated with cerebral oedema, which might be reversible and a treatable factor. Oedema is the leading cause of death in acute hepatic failure. It may respond to the administration of mannitol and measures to control agitation.[61] For patients with acute hepatic failure who have significant hepatic encephalopathy, liver transplantation increases survival from 20 to 80 per cent, making rapid and accurate diagnosis vital. Survival of

liver transplant patients with neuropsychiatric involvement is significantly lower if there is liver disease alone.[61] There is also an increased incidence of liver disease among patients with primary psychiatric disorders, including substance use disorders.[62] B virus carriers are almost three times more likely to have psychiatric disorders than comparison subjects.[63]

Endocrine disorders

Diabetes mellitus

Type I diabetes (insulin-dependent diabetes mellitus) occurs when the pancreas' ability to secrete insulin is clinically impaired. Hyperglycaemic symptoms emerge when 80 to 90 per cent of islet cells fail to produce insulin. Around 90 per cent of diabetics have type II diabetes (non-insulin-dependent diabetes mellitus). Type II diabetes is characterized by peripheral resistance to the action of insulin and decreased insulin secretion, in spite of the presence of elevated serum glucose levels. Patients with type II diabetes can often avoid or postpone the need for insulin treatment with appropriate diet and exercise. Both type I and II diabetes are associated with a genetic predisposition.[64]

The most frequent psychiatric disorders in patients with diabetes are anxiety and depressive disorders. Among general populations of diabetics, anxiety disorders occurred in up to 45 per cent and depressive disorders in up to 33 per cent.[65] Rosenthal *et al.*[66] in a 3-year prospective study of hospitalizations and mortality in older patients with diabetes, found that the combined presence of retinopathy and a high depression score on the Geriatric Depression scale had the strongest relationship with mortality. Patients with diabetes are twice as likely to experience depression as those without diabetes; this holds true for both type I and II diabetes.[66] Patients with schizophrenia are at increased risk for developing type II diabetes[67] is growing interest in the possibility that there are shared inherited risk factors for the two disorders,[68] though the evidence is weak and largely circumstantial. In addition, second generation antipsychotic medications, especially olanzapine, are associated with type II diabetes mellitus and abnormal glucose metabolism.

Diabetic patients who have psychiatric disorders can have less disease morbidity when their psychiatric disorders are appropriately treated, highlighting the importance of monitoring diabetic patients for psychiatric disorders and monitoring psychiatric patients for excess weight and diabetes.[69] Independent of the level of physical illness present in type I or II diabetes, the presence of anxiety and/or depression is important in determining the quality of a patient's life.[70] Treatment adherence problems complicate care, particularly in children and adolescents with type I diabetes. A great deal of patience, family support, and education is necessary to minimize passive and active non-compliance.

Hypothyroidism

Hypothyroidism is usually the result of primary failure or ablation of the thyroid gland, hypothalamic dysfunction, pituitary dysfunction, autoimmune thyroiditis, or lithium therapy. Clinical manifestations of hypothyroidism include fatigue, cold intolerance, lethargy, weakness, weight gain, constipation, menstrual irregularities, hair loss, slow reaction time, oedema, delayed reflexes, and bradycardia. Hypothyroidism occurs in as many as 10 per cent of patients taking lithium; lithium-induced hypothyroidism is more likely to occur in women.[70]

The association between clinical hypothyroidism and depression is well known. Gold *et al.*[71] found that 5 per cent of a series of 250 patients with major depressive syndromes had at least subclinical hypothyroidism. In many patients with hypothyroidism, the depression responds to thyroid hormone replacement alone,[72] but the response may take a long time. When that is the case, antidepressants are indicated and efficacious.[71]

Hyperthyroidism

The most frequent clinical manifestations of hyperthyroidism are nervousness, diaphoresis, hypersensitivity to heat, palpitations, fatigue, weight loss, tachycardia, dyspnoea, and weakness. The most common causes include Graves' disease, toxic adenoma, and toxic multinodular goitre. Less common causes include Hashimoto's thyroiditis, postpartum hyperthyroidism, and factitious hyperthyroid state.

As with hypothyroidism, depressive and anxiety syndromes are the most common psychiatric conditions seen among patients with hyperthyroid states; there is a three-fold increased risk for development of mood disorder following hospitalization with hyperthyroidism.[73,74] When patients have depressive or anxiety syndromes in the context of hyperthyroidism, and have no past histories of psychiatric disorders, the psychiatric symptoms resolve more than 90 per cent of the time when the hyperthyroidism resolves. This obviates the need for other psychiatric interventions unless antithyroid medication, radioactive iodine, or thyroid surgery has not been successful.[74] Anxiety symptoms will disappear in direct relation to the reduction of thyroid hormone levels. Depressive symptoms are not quite so linearly related and may resolve at a slower pace as thyroid hormone level normalize.

Hypoparathyroidism

Parathyroid hormone mobilizes calcium from bone, induces renal reabsorption of calcium, increases renal clearance of inorganic phosphate, and promotes intestinal reabsorption of calcium. Hypoparathyroidism would be expected to result in hypocalcaemia, which can cause delirium. Hypoparathyroidism can result from autoimmune destruction of the parathyroid glands, removal of the parathyroids, disruption of the glands' blood supply, tumour, or neck irradiation. Medical and neuropsychiatric symptoms and signs are related to the level of serum calcium and the rate at which hypocalcaemia develops. Faster the hypocalcaemia develops, the more likely delirium and other neuropsychiatric symptoms are to occur. The most frequent symptoms are caused by neuromuscular irritability and include paraesthesias, carpal pedal spasm, laryngospasm, blepharospasm, and bronchospasm.[75] Cardiovascular manifestations include prolonged Q–T interval, heart block, and congestive heart failure. The most common neuropsychiatric symptoms and signs are seizures, EEG abnormalities, increased intracranial pressure, disorientation, confusion, and extrapyramidal symptoms. The mainstays of treatment are calcium and vitamin D. Neuropsychiatric syndromes should resolve with the normalization of serum calcium.

Hyperparathyroidism

Typically, hypercalcaemia is discovered by routine laboratory testing in patients without obvious illness. Primary hyperparathyroidism is the most common cause of hypercalcaemia among adult patients; among hospitalized patients, malignancy is the most

common cause.[75] Reversible hyperparathyroidism and hypercalcaemia are also associated with lithium therapy.[76] In primary hyperparathyroidism, parathyroid hormone is secreted inappropriately, despite an elevation in the ionized calcium level. Signs and symptoms of hyperparathyroidism include nausea, vomiting, anorexia, constipation, proximal muscle weakness, polyuria, polydipsia, impaired renal function, hypertension, short Q–T interval, bradycardia, and a number of neuropsychiatric symptoms. The latter include lethargy, drowsiness, impaired concentration ability, and confusion. In severe cases, there may be stupor or coma, psychosis, and cognitive impairment are common in patients who have serum calcium levels above 16 mg/dl. Depressive symptoms, but not cognitive symptoms, tend to resolve with treatment.[77] Cognitive symptoms may improve, but residual symptoms usually remain.

Cushing's syndrome

Hypersecretion of cortisol by the adrenal gland can result in Cushing's syndrome. Cushing's syndrome can also be due to exogenous ACTH or glucocorticoid administration, or endogenous hyperproduction of these hormones. Because a physiological release of cortisol occurs during periods of stress or duress, it is common to see elevations of serum cortisol during the courses of many psychiatric disorders, including major depressive disorder, alcoholism, anorexia nervosa, panic disorder, and psychoactive substance-withdrawal syndromes. The more common clinical signs and symptoms of Cushing's syndrome, whether endogenous or exogenous, include fat redistribution, menstrual irregularities, dysphoria, thin skin, moon facies, increased appetite, sleep disturbances, hypertension, hypercholesterolaemia, hypertriglyceridaemia, poor concentration, impaired memory, euphoria, glucose intolerance, striae, and hirsutism.

At least half of all patients with Cushing's syndrome will experience depressive or manic symptoms[78]; the symptoms will be moderate to severe in half of these patients. Many will also experience psychotic symptoms. Symptoms are dose-related when due to exogenous steroids. Depression or mania due to Cushing's syndrome will eventually remit when the hypercortisolaemia is corrected, the return to euthymia is usually gradual. When depression or mania is slow to remit, treatment with antidepressants or mood stabilizers is warranted.

Addison's disease

Addison's disease is the result of an autoimmune process that destroys the adrenal glands; it is the most common cause of primary adrenal insufficiency in the industrialized world, accounting for about 65 per cent of cases.[70] Both glucocorticoid and mineralocorticoid secretion are diminished in this condition. Clinical manifestations of adrenal insufficiency include weight loss, fatigue, vomiting, diarrhoea, anorexia, and salt-craving. Patients with Addison's disease will require lifelong replacement of both glucocorticoids and mineralocorticoids. Patients may be misdiagnosed with major depressive disorder, personality disorder, dementia, or somatoform disorders. It is not uncommon for the diagnosis to be delayed for many months; Addison's disease is a disorder to be continually mindful about in a patient with treatment-resistant depression.

Hyperprolactinaemia

Prolactin is synthesized in the pituitary gland; its secretion is increased during pregnancy, enhancing breast development. Prolactin secretion is inhibited by glucocorticoids and thyroid hormone, it is predominantly under the inhibitory control of dopamine.[70] Dopamine is, in fact, prolactin inhibiting factor. This is why dopamine blocking medications such as neuroleptics can cause hyperprolactinaemia. Symptoms of hyperprolactinaemia include breast development and lactation. These bothersome side effects of neuroleptics, particularly in men, can be modulated by changing to a more favourable neuroleptic medication, such as a second generation antipsychotic.

Hypopituitarism

Hypopituitarism occurs when multiple pituitary hormones exhibit decreased secretion. It can be due to either gland destruction or inadequate stimulation by factors that regulate pituitary functioning. Common causes of hypopituitarism are pituitary adenomas, hypothalamic tumours, metastatic carcinoma (especially breast and bronchus), and cerebral trauma and haemorrhage.[70] Other causes include vascular disorders, immunological conditions, and a variety of congenital anomalies. Clinical manifestations depend on which hormones are deficient; signs and symptoms are those of the individual deficiency states. Other signs that may be present when there is hypopituitarism include headache, visual loss, and radiographically discovered sella enlargement. Treatment of patients with hypopituitarism involves hormone replacement therapy and surgery when accessible lesions are present.[70]

Autoimmune disorders

Systemic lupus erythematosus

Systemic lupus erythematosus is characterized by the production of autoantibodies that injure tissue in several organ systems, most frequently the skin, central nervous system, kidney, and lungs. In addition, non-specific symptoms occur in a large majority of patients, including fatigue, fever, weight loss, arthralgias, and myalgias.[79] The most common presenting symptoms are fever, arthralgias, butterfly rash, photosensitivity, Raynaud's phenomenon, and mucous ulcers. The laboratory diagnostic hallmark of systemic lupus erythematosus is the production of high-titre autoantibodies directed against a variety of cell nucleus components (antinuclear antibodies). Systemic lupus erythematosus has no known cure, so treatment is based on symptom relief, suppression of inflammation, and preventing future pathology.

The most common psychiatric presentations of active systemic lupus erythematosus are psychosis, delirium, seizures, and cognitive dysfunction.[80] Autoantibodies to neuronal membranes, which interfere with the ability of neurones to respond to stimuli, may account for most neuropsychiatric deficits and symptoms,[81] though CNS vasculitis during the course of this disease may also play a role in some patients. Symptoms in lupus patients are likely to be related to unique sets of autoantigens,[82] and are related to antibodies against N-methyl-D-aspartate (NMDA) receptors.[83] Antibodies may be partially responsible for depression, short-term memory problems, and new learning difficulties in lupus patients. Significantly associated with declining cognitive function are consistently positive antiphospholipid antibodies, consistent steroid use, diabetes, and higher depression scores.[83] Mood syndromes are probably the most common psychiatric presentation of patients with systemic lupus erythematosus,[84] but mood change is not

always due to involvement of the central nervous system. Patients are frequently treated with steroids, which raise the possibility of steroid-induced neuropsychiatric syndromes. However, in many cases, addition of steroids to treatment may improve psychiatric syndromes.

Renal disorders

Acute renal failure

Acute renal failure is an abrupt decrease in renal function sufficient to result in azotaemia—retention of nitrogenous waste in the body.[85] Acute renal failure can result from a decrease of renal blood flow (prerenal azotaemia), intrinsic renal disease (renal azotaemia), or obstruction of urine flow (postrenal azotaemia). Prerenal azotaemia can be caused by renal arterial occlusion or a decrease in the effective blood volume (e.g. haemorrhage, congestive heart failure, diarrhoea). Intrinsic renal azotaemia is most commonly caused by acute tubular necrosis due to an acute ischaemic or nephrotoxic insult. Azotaemia is due to obstruction of the urine collecting system; this may occur when there is bladder outlet obstruction or ureteral obstruction.

Medical complications of acute renal failure include hyperkalaemia, hyperuricaemia, arrhythmias, anaemia, coagulopathies, vomiting, nausea, and urinary tract infections. Metabolic perturbations can lead to delirium. Neuropsychiatric manifestations include somnolence, asterixis (flapping tremor), neuromuscular irritability, and seizures. Mental status abnormalities in acute (but not chronic) renal failure begin to occur for most adults when the serum creatinine level acutely rises to about 4.0 mg/dl. In oliguric renal failure, serum blood urea nitrogen levels can be expected to rise by about 10 to 20 mg/dl per day. Serum creatinine levels can be expected to rise by about 1 mg/dl per day. Neuropsychiatric complications of acute renal failure are best treated by correcting the underlying cause of the renal failure. Dialysis may be used to manage acute manifestations. While awaiting reversal of neuropsychiatric manifestations, symptomatic management with antiseizure medications and neuroleptics may be necessary.

Chronic renal failure and end stage renal disease

Chronic renal failure is a progressive and irreversible loss of renal function.[86] The most common aetiologies of renal insufficiency ultimately leading to end stage renal disease are diabetes, hypertension, and glomerulonephritis. Loss of up to 75 per cent of glomerular filtration rate does not usually result in pronounced clinical symptoms, as the remaining glomeruli adapt with hyperfiltration. Serum creatinine is a sensitive indicator of early, subclinical, chronic renal failure. For example, the doubling of serum creatinine from 0.7 to 1.4 mg/dl signifies a loss of approximately 50 per cent of glomerular filtration rate, emphasizing the importance of early detection and prevention.

Patients with chronic renal failure usually become symptomatic when glomerular filtration rate is less than 10 ml/min. Uraemia affects every organ system, including the central nervous system. Neuropsychiatric manifestations of chronic renal failure include irritability, insomnia, lethargy, anorexia, seizures, and restless legs syndrome.[86] In contrast to acute renal failure—where neuropsychiatric signs and symptoms may appear with a creatinine level as low as 4 mg/dl—in chronic renal failure, patients may have a normal mental status examination with a serum creatinine level as high as 9 to 10 mg/dl. Symptomatic treatments with low-dose neuroleptics, antiseizure medications, or benzodiazepines are sometimes necessary in chronic renal failure.

Kimmel *et al.*[86] studied the prevalence of hospitalizations for psychiatric illness in patients with end stage renal disease and compared that rate with four other chronic medical conditions (diabetes, ischaemic heart disease, cerebrovascular disease, and peptic ulcer disease). Hospitalizations for mental disorders were 1.5 to 3.0 times higher in these patients than in patients with the four other chronic diseases. Dementias and depression were the most common reasons for hospitalization. With end stage renal disease are almost twice as likely to die by suicides as the general populations, after controlling for other potential demographic and clinical contributors to suicide risk.[87]

Sexual dysfunction is very common in patients with end stage renal disease. Abrams *et al.*[88] found that 75 per cent of his sample of men with this disease reported a decrease in frequency of sexual intercourse of at least 50 per cent. Disruptions in sexual function, which may be physiological (e.g. vascular complications of diabetes, fatigue following dialysis treatments) or psychological or both, account for at least a portion of the dysphoria experienced by patients with end stage renal disease.

The definitive treatments for most patients with chronic renal failure are transplantation or haemodialysis. In general, transplantation is encouraged because of a better quality of life and a greater chance for rehabilitation and symptom resolution. Researchers in three separate prospective studies found that patients who received renal transplants experienced better physical and psychological outcomes than patients who remained on dialysis.[89] Neuropsychiatric signs and symptoms resolve much more completely with transplantation than with haemodialysis. Psychiatric aspects of organ transplantation are discussed later. The psychiatric aspects of haemodialysis are discussed next.

Haemodialysis

The average patient on haemodialysis requires 3.5 h of dialysis three times per week to achieve adequate creatinine clearance.[86] Haemodialysis has enabled the survival of countless thousands of patients with chronic renal failure and provides a temporary management tool for patients on transplantation waiting lists. However, it is not a benign procedure, and has a number of potential neuropsychiatric complications. Patients on haemodialysis are at high risk for developing volume overload, pulmonary oedema, hyperkalaemia, hyperphosphataemia, and metabolic bone disease if compliance with restricted diet and fluid intake is not optimal. Patient adherence to these diet and fluid-intake protocols are used as one of the criteria for making decisions about appropriateness for transplantation. Psychiatric reasons for non-adherence should be addressed and are usually reversible, with the exception of personality disorders. These include mood disorders, phobias, panic disorder, substance-related disorders, adjustment disorder, and cognitive disorders.

Haematological disorders

Anaemia due to vitamin deficiency

Both folic acid and cobalamin (vitamin B12) are necessary for the production of DNA; in their absence, the nucleus of the cell cannot

undergo normal mitosis. The main cause of folic acid deficiency is dietary insufficiency.(90) This commonly occurs in severe alcoholics. The main cause of cobalamin deficiency is malabsorption. The major clinical manifestations of folate or vitamin B12 deficiency include fatigue, pallor, and for cobalamin deficiency, neuropsychiatric manifestations. The latter include loss of proprioception in the lower extremities, loss of vibratory perception, anosmial, forgetfulness, and even dementia. Diagnosis is based on measurement of serum levels of vitamin B12 and folate. Treatment is replacement of folate (1 mg/day or improved diet) and vitamin B12 (administered parenterally).

In cobalamin deficiency, neuropsychiatric findings can occur even when megaloblasts and anaemia are absent.(91) Patients with cobalamin levels between 100 and 200 pg/ml, and especially those with levels less than 100 pg/ml, may have cobalamin-reversible neuropsychiatric deficits.(138) Even dementia due to chronic cobalamin deficiency may be partially reversible when diagnosed and aggressively treated. It is also associated with a higher risk of depression, especially in older adults; anaemia severity and depression severity scores are highly correlated.(92) Treatment of the anaemia may result in resolution of depressive symptoms; in some cases depressive symptoms are slower to resolve and may require extended treatment with antidepressant medication, particularly for patients predisposed to mood disorders.

Iron-deficiency anaemia

When erythrocyte-cytoplasm production is abnormally low due to the reduced production or availability of one of the three components of haemoglobin (iron, globin, or haem), the ratio of cytoplasm to the contents of the rest of the cell declines. This results in a microcytic anaemia. In most cases, microcytic anaemia is due to iron-deficient haemopoiesis.(92) Deficiency is usually due to an iron-poor diet or defective iron utilization by the body. Clinical manifestations of iron-deficiency anaemia are related to the severity of the iron deficiency and include severe fatigue, pallor, changes in nail curvature ('spoon nails'), and, at times, pica and cheilosis at the corners of the mouth. A diminished haematocrit and mean corpuscular volume, and low serum iron are the cornerstones of the diagnosis. The severe fatigue associated with iron deficiency can be misdiagnosed as major depressive disorder; treatment with an antidepressant will not help unless there is also clinical major depression apart from the iron deficiency. Iron replenishment to correct the underlying deficiency is the specific. With iron replenishment, the haemoglobin should correct to normal, and symptoms should resolve, within 4 to 6 weeks.

Conclusions

There is increasing recognition that patients with psychiatric signs and symptoms frequently have associated medical disorders. Interactions between the disorders can be quite complex and may involve neuropsychiatric manifestations of medical illness, medical effects of psychiatric treatments, psychiatric effects of medical treatment, increased medical illness related to factors inherent in the psychiatric condition, and maladaptive personality styles or disorders that affect outcomes of medical and psychiatric illness. Psychiatric disorders in patients with other medical illnesses increase mortality, morbidity, health care costs, and decrease adherence to medical therapies. Appropriate and cost-effective management must integrate knowledge from psychiatry and primary care to assure maximal therapeutic gains for the patient. Those involved in consultation to primary care providers must understand both the basic pathophysiology and clinical characteristics of the medical disorders and their treatments, as well as possess the psychiatric knowledge necessary to ensure that the patient with medical illness is adequately managed.

Further information

Robinson, M.J. and Owen, J.A. (2005). Psychopharmacology. In *Textbook of psychosomatic medicine* (ed. J.L. Levenson), pp. 871–3. American Psychiatric Publishing, Inc., Washington, DC.

Saha, S., Chant, D., and McGrath, J. (2007). A systematic review of mortality in schizophrenia. *Archives of General Psychiatry*, **64**, 1123–31.

Schmitz, N., Thefeld, W., and Kruse, J. (2006). Mental disorders and hypertension: factors associated with awareness and treatment of hypertension in the general population of Germany. *Psychosomatic Medicine*, **68**, 246–52.

References

1. Chapman, D.P., Perry, G.S., and Strine, T.W. (2005). The vital link between chronic disease and depressive disorder. *Preventing Chronic Disease*, **2**, 1–12.
2. Barsky, A.J., Orav, J., and Bates, D.W. (2005). Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity. *Archives of General Psychiatry*, **62**, 903–10.
3. Brod, J. (1970). Circulatory changes underlying blood pressure elevation during acute emotional stress (mental arithmetic) in normotensive and hypertensive subjects. *Clinical Scientist*, **18**, 269–78.
4. Torian, L., Davidson, E., Fulop, G., *et al.* (1992). The effect of dementia on acute care in a geriatric medical unit. *International Psychogeriatrics*, **4**, 231–9.
5. Thomas, S.A., Friedmann, E., and Kelley, F.J. (2001). Living with an implantable cardioverter-defibrillator: a review of the current literature related to psychosocial factors. *The Journal of Cardiovascular Nursing*, **12**, 156–63.
6. Lown, B., DeSilva, R.A., and Reich, P. (1980). Psychophysiologic factors in sudden cardiac death. *The American Journal of Psychiatry*, **137**, 1325–35.
7. Tennant, C. (1999). Life stress, social support and coronary heart disease. *The Australian and New Zealand Journal of Psychiatry*, **33**, 636.
8. Frasure-Smith, N., Lesperance, F., and Talajic, M. (1995). Depression and 18-month prognosis after myocardial infarction. *Circulation*, **91**, 999–1005.
9. Benson, H., Alexander, S., and Feldman, C.L. (1975). Decreased premature ventricular contractions through the use of relaxation response in patients with stable ischaemic heart disease. *Lancet*, **2**, 380–2.
10. Grippo, A.J., Santos, C.M., Johnson, R.F., *et al.* (2004). Increased susceptibility to ventricular arrhythmias in a rodent model of experimental depression. *American Journal of Physiology—Heart and Circulatory Physiology*, **286**, H619–26.
11. Musselman, D.L., Evans, D.L., and Nemeroff, C.B. (1998). The relationship of depression to cardiovascular disease: epidemiology, biology, and treatment. *Archives of General Psychiatry*, **55**, 580–92.
12. Brod, J. (1970). Circulatory changes underlying blood pressure elevation during acute emotional stress (mental arithmetic) in normotensive and hypertensive subjects. *Clinical Scientist*, **18**, 269–78.
13. Harburg, E., Erfurt, J.C., and Hauenstein, L.S. (1973). Socio-ecological stress, suppressed hostility, skin color, and black-white male blood pressure: Detroit. *Psychosomatic Medicine*, **35**, 276–96.
14. Bankier, B. and Littman, A.B. (2002). Psychiatric disorders and coronary heart disease in women—a still neglected topic: review of the literature from 1971 to 2000. *Psychotherapy and Psychosomatics*, **71**, 133–40.

15. Rozanski, A., Bairey, C.N., and Krantz, D.S. (1988). Mental stress and the induction of silent myocardial ischemia in patients with coronary artery disease. *The New England Journal of Medicine*, **318**, 1005–11.
16. Ruberman, W., Weinblatt, A.B., Goldberg, J.D., *et al.* (1984). Psychosocial influences on mortality after myocardial infarction. *The New England Journal of Medicine*, **311**, 552–9.
17. Dickens, C.M., McGowan, L., Percival, C., *et al.* (2004). Lack of a close confidant, but not depression, predicts further cardiac events after myocardial infarction. *Heart*, **90**, 518–22.
18. Mayou, R.A., Gill, D., Thompson, D.R., *et al.* (2000). Depression and anxiety as predictors of outcome after myocardial infarction. *Psychosomatic Medicine*, **62**, 212–19.
19. Lesperance, F., Frasure-Smith, N., and Talajic, M. (1996). Major depression before and after myocardial infarction: its nature and consequences. *Psychosomatic Medicine*, **58**, 99–110.
20. Carney, R.M., Rich, M.W., and Freedland, K.E. (1988). Major depressive disorder predicts cardiac events in patients with coronary artery disease. *Psychosomatic Medicine*, **50**, 627–33.
21. Plotsky, P.M., Owens, M.J., and Nemeroff, C.B. (1998). Psychoneuroendocrinology of depression the Hypothalamic-Pituitary-Adrenal Axis. *The Psychiatric Clinics of North America*, **21**, 293–307.
22. Musselman, D.L., Evans, D.L., and Nemeroff, C.B. (1998). The relationship of depression to cardiovascular disease. *Archives of General Psychiatry*, **55**, 580–92.
23. Langosch, W., Seer, P., Brodner, G., *et al.* (1982). Behavior therapy with coronary heart disease patients: results of a comparative study. *Journal of Psychosomatic Research*, **26**, 475–84.
24. Preckel, D., von Kanel, R., Kudielka, B.M., *et al.* (2005). Over commitment to work is associated with vital exhaustion. *International Archives of Occupational and Environmental Health*, **78**, 117–22.
25. Friedman, M. and Thoresen, C.E. (1986). Alteration of type A behavior and its effect on cardiac recurrences in post-myocardial infarction patients: summary results of the recurrent coronary prevention project. *American Heart Journal*, **112**, 653–65.
26. Sher, L. (2005). Type D personality: the heart, stress, and cortisol. *Quarterly Journal of Medicine*, **98**, 323–9.
27. Denollet, J., Pedersen, S.S., Vrints, C.J., *et al.* (2006). Usefulness of type D personality in predicting five-year cardiac events above and beyond concurrent symptoms of stress in patients with coronary heart disease. *The American Journal of Cardiology*, **97**, 970–3.
28. Katon, W.J., Richardson, L., Lozano, P., *et al.* (2004). The relationship of asthma and anxiety disorders. *Psychosomatic Medicine*, **66**, 349–55.
29. Ten Brinke, A., Sterk, P.J., Masclee, A.A.M., *et al.* (2005). Risk factors of frequent exacerbations in difficult-to-treat asthma. *The European Respiratory Journal*, **26**, 812–18.
30. Greenberg, D.B., Halperin, P., Kradin, R.L., *et al.* (2000). Internal medicine and medical subspecialties. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), pp. 548–608. American Psychiatric Publishing, Washington, DC.
31. Norwood, R. (2006). Prevalence and impacts of depression in chronic obstructive pulmonary disease. *Current Opinion in Pulmonary Medicine*, **12**, 113–17.
32. Porzelius, J., Vest, M., and Nochomovitz, M. (1992). Respiratory function, cognitions, and panic in chronic obstructive pulmonary patients. *Behavioural Research and Therapy*, **30**, 75–7.
33. Strudsholm, U., Johannessen, L., Foldager, L., *et al.* (2005). Increased risk for pulmonary embolism in patients with bipolar disorder. *Bipolar Disorders*, **7**, 77–81.
34. Joukamaa, M., Heliovaara, M., Knekt, P., *et al.* (2006). Schizophrenia, neuroleptic medication, and mortality. *The British Journal of Psychiatry*, **188**, 122–7.
35. Clouse, R.E. and Lustman, P.J. (1989). Value of recent psychological symptoms in identifying patients with esophageal contraction abnormalities. *Psychosomatic Medicine*, **51**, 570–6.
36. Van Handel, D. and Fass, R. (2005). The pathophysiology of non-cardiac chest pain. *Journal of Gastroenterology and Hepatology*, **20**, S6–11.
37. Clouse, R.E. and Lustman, P.J. (1983). Psychiatric illness and contraction abnormalities of the esophagus. *The New England Journal of Medicine*, **309**, 1337–42.
38. Castell, D.O. and Richter, J.E. (1987). Edrophonium testing for esophageal pain: concurrence and discord. *Digestive Disease Scientist*, **32**, 897–9.
39. Cybulska, E.M. (1997). Globus hystericus: a somatic symptom of depression? *Psychosomatic Medicine*, **59**, 67–9.
40. Drossman, D.A., Sandler, R.S., and McKee, D.C. (1982). Bowel patterns among subjects not seeking health care. *Gastroenterology*, **83**, 529–34.
41. Bhatia, V. and Tandon, R.K. (2005). Stress and the gastrointestinal tract. *Journal of Gastroenterology and Hepatology*, **20**, 332–44.
42. Clouse, R.E., Richter, J.E., Heading, R.C., *et al.* Functional esophageal disorders. *Gut*, **45**, 31–6.
43. Clouse, R.E. and Lustman, P.J. (2005). Use of psychopharmacological agents for functional gastrointestinal disorders. *Gut*, **54**, 1332–41.
44. Walker, E.A., Gelfand, M.D., Gelfand, A.N., *et al.* (1996). The relationship of current psychiatric disorders to functional disability and distress in patients with inflammatory bowel disease. *General Hospital Psychiatry*, **18**, 220–9.
45. Bennett, P. and Wilkinson, S. (1985). A comparison of psychological and medical treatment of the irritable bowel syndrome. *The British Journal of Clinical Psychology*, **24**, 215–16.
46. Maunder, R. (2005). Evidence that stress contributes to inflammatory bowel disease: evaluation, synthesis, and future directions. *Inflammatory Bowel Diseases*, **11**, 600–8.
47. Mawdsley, J.E. and Rampton, D.S. (2005). Psychological stressing IBD: new insights into pathogenic and therapeutic implications. *Gut*, **54**, 1481–91.
48. Whorell, P.J., Prior, A., and Faragher, E.B. (1984). Controlled trial of hypnotherapy in the treatment of severe refractory irritable bowel syndrome. *Lancet*, **2**, 1232–4.
49. Cremonini, F., Wise, J., Moayyedi, P., *et al.* (2005). Diagnostic and therapeutic use of proton pump inhibitors in non-cardiac chest pain: a meta-analysis. *The American Journal of Gastroenterology*, **100**, 1226–9.
50. Littman, A.B. and Ketterer, M.W. (2000). Behavioural medicine. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), pp. 1080–95. American Psychiatric Publishing, Washington, DC.
51. Magni, G. (1987). On the relationship between chronic pain and depression when there is no organic lesion. *Pain*, **31**, 1–21.
52. Whitehead, W.E. (1992). Behavioural medicine approaches to gastrointestinal disorders. *Journal of Consulting and Clinical Psychology*, **60**, 605–12.
53. Ang, T.L., Fock, K.M., Ng, T.M., *et al.* (2005). A comparison of the clinical, demographic and psychiatric profiles among patients with erosive and non-erosive reflux disease in a multi-ethnic Asian country. *World Journal of Gastroenterology*, **11**, 3558–61.
54. Elfhag, K. and Rossner, S. (2005). Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain. *Obesity Reviews*, **6**, 67–86.
55. Morriss, R. and Mohammed, F.A. (2005). Metabolism, lifestyle and bipolar disorder. *Journal of Psychopharmacology*, **19**, 94–101.
56. Casey, D.E. (2005). Metabolic issues and cardiovascular disease in patients with psychiatric disorders. *The American Journal of Medicine*, **118**(Suppl. 2), 15S–22S.
57. Prashanth, L.K., Sinha, T.S., Ravishankar, S., *et al.* (2005). Prognostic factors in patients presenting with severe neurological forms of Wilson's disease. *Quarterly Journal of Medicine*, **98**, 557–63.
58. Dening, D.C. and Berrios, G.E. (1989). Wilson's disease: psychiatric symptoms in 195 cases. *Archives of General Psychiatry*, **46**, 1126–34.
59. National Cholesterol Education Program. (1993). Summary of the second report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation and treatment of high blood

cholesterol in adults. (Adult Treatment Panel 11). *The Journal of the American Medical Association*, **264**, 3015–23.
60. Freedman, D.S., Byers, T., Barrett, D.H., *et al.* (1995). Plasma lipid levels and psychological characteristic in men. *American Journal of Epidemiology*, **141**, 507–17.
61. Medici, V., Mirante, V.G., Fassati, L.R., *et al.* (2005). Liver transplantation for Wilson's disease: the burden of neurological and psychiatric disorders. *Liver Transplantation*, **11**, 1056–63.
62. Crone, C.C., Gabriel, G.M., and DiMartini, A. (2006). An overview of psychiatric issues in liver disease for the consultation-liaison psychiatrist. *Psychosomatics*, **47**, 188–205.
63. Atesci, F.C., Cetin, B.C., Oguzhanoglu, N.K., *et al.* (2005). Psychiatric disorders and functioning in Hepatitis B virus carriers. *Psychosomatics*, **46**, 142–7.
64. Lustman, P.J., Griffith, L.S., Clouse, R.E., *et al.* (1986). Psychiatric illness in diabetes mellitus: relationship to symptoms and glucose control. *The Journal of Nervous and Mental Disease*, **174**, 736–42.
65. Kathol, R. (2000). In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), pp. 579–84. American Psychiatric Publishing, Washington, DC.
66. Rosenthal, M.J., Morley, J.E., Fajardo, M., *et al.* (1998). Hospitalization and mortality of diabetes in older adults. *Diabetes Care*, **21**, 231–5.
67. Chafetz, L., White, M.C., Collins-Bride, G., *et al.* (2005). The poor general health of the severely mentally ill: impact of schizophrenic diagnosis. *Community Mental Health Journal*, **41**, 169–84.
68. Gough, S.C.L. and O'Donovan, M.C. (2005). Clustering of metabolic comorbidity in schizophrenia: a genetic contribution? *Journal of Psychopharmacology*, **19**, 47–55.
69. McIntyre, R.S., Mancini, D.A., Pearce, M.M., *et al.* (2005). Mood and psychotic disorders and type 2 diabetes: a metabolic trial. *Canadian Journal of Diabetes*, **29**, 122–32.
70. Goldman, M.B. (2002). Neuropsychiatric features of endocrine disorders. In *The American psychiatric publishing textbook of neuropsychiatry* (4th edn) (eds. S. Yudofsky and R. Hales), pp. 519–40. American Psychiatric Publishing, Washington, DC.
71. Gold, M.S., Pottash, A.L.C., and Extein, I. (1981). Hypothyroidism and depression. *The Journal of the American Medical Association*, **245**, 1919–22.
72. Rouchell, A.M., Pounds, R., and Tierney, J.G. (2000). Depression. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), pp. 310–45. American Psychiatric Publishing, Washington, DC.
73. Williams, R.H. (1946). Thiouracil treatment of thyrotoxicosis. *The Journal of Clinical Endocrinology and Metabolism*, **6**, 1–22.
74. Thomsen, A.F., Kvist, T.K., Andersen, P.K., *et al.* (2005). Increased risk of affective disorder following hospitalization with hyperthyroidism—a register based study. *European Journal of Endocrinology*, **152**, 535–43.
75. Brown, G.G., Preisman, R.C., and Kleerekoper, M. (1987). Neurobehavioural symptoms in mild primary hyperparathyroidism: related to hypercalcemia but not improved by parathyroidectomy. *Henry Ford Hospital Medical Journal*, **35**, 211–15.
76. Khandwala, H.M. and Van Uum, S. (2006). Reversible hypercalcemia and hyperparathyroidism associated with lithium therapy: case report and review of the literature. *Endocrine Practice*, **12**, 54–8.
77. Allerheiligen, D.A., Schoeber, J., Houston, R.E., *et al.* (1998). Hyperparathyroidism. *American Family Physician*, **57**, 1795–802.
78. Thompson, J.M., Gallagher, P., Hughes, J.H., *et al.* (2005). Neurocognitive impairment in euthymic patients with bipolar affective disorder. *The British Journal of Psychiatry*, **186**, 32–40.
79. Yagniak, P.M. and Cohen, M.M. (1988). Systemic lupus erythematosus nervous system involvement. In *Diagnosis and management of rheumatic diseases* (2nd edn) (ed. W. Katz), pp. 220–3. J.B. Lippincott, Philadelphia, PA.
80. Greenberg, D.B. (2000). Systemic lupus erythematosus. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), p. 585. American Psychiatric Publishing, Washington, DC.
81. Baker, M. (1973). Psychopathology in SLE: psychiatric observations. *Seminars in Arthritis and Rheumatology*, **3**, 95–110.
82. Margutti, P., Sorice, M., Conti, F., *et al.* (2005). Screening of an endothelial cDNA library identifies the C-terminal region of Nedd5 as a novel autoantigen in systemic lupus erythematosus with psychiatric manifestations. *Arthritis Research & Therapy*, **7**, R896–903.
83. Omdal, R., Brokstad, K., Waterloo, K., *et al.* (2005). Neuropsychiatric disturbances in systemic lupus erythematosus are associated with antibodies against NMDA receptors. *European Journal of Neurology*, **12**, 392–400.
84. McLaurin, E.Y., Holliday, S.L., Williams, P., *et al.* (2005). Predictors of cognitive dysfunction in patients with systemic lupus erythematosus. *Neurology*, **64**, 297–303.
85. Cohen, L.M. (2000). Renal disease. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), pp. 573–8. American Psychiatric Publishing, Washington, DC.
86. Kimmel, P.L., Thamer, M., Richard, C.M., *et al.* (1998). Psychiatric illness in patients with end-stage renal disease. *The American Journal of Medicine*, **105**, 214–21.
87. Kurella, M., Kimmel, P.L., Young, B.S., *et al.* (2005). Suicide in the United States end-stage renal disease program. *Journal of the American Society of Nephrology*, **16**, 774–81.
88. Abrams, H.S., Hester, L.R., and Sheridan, W.F. (1975). Sexual functioning in patients with chronic renal failure. *The Journal of Nervous and Mental Diseases*, **160**, 220–6.
89. Russell, J.D., Beecroft, M.L., Ludlow, D., *et al.* (1992). The quality of life in renal transplantation—a prospective study. *Transplantation*, **54**, 656–60.
90. Stabler, S.P. and Allen, R.H. (1990). Clinical spectrum and diagnosis of cobalamin deficiency. *Blood*, **76**, 871–81.
91. Greenberg, D.B. (2000). Cobalamin deficiency. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), p. 586. American Psychiatric Publishing, Washington, DC.
92. Onder, G., Penninx, B.W., Cesari, M., *et al.* (2005). Anemia is correlated with depression in older adults: results from the InCHIANTI study. *Gerontology*, **60**, 1168–72.
93. Badminton, M.N. and Elder, G.H. (2005). *Journal of Inherited Metabolic Disease*, **28**, 277–86.

5.3.5 Psychiatric aspects of infections

José Luis Ayuso-Mateos

Neuropsychiatric disturbances stemming from infectious diseases are widespread in both the industrialized world and developing countries. Such neuropsychiatric syndromes are not necessarily the result of infectious processes directly involving the central nervous system, they may also be complications of systemic infections. There are many microbial, viral, and parasitic agents, as well as other types of infectious substances, which can affect the central nervous system, leading to the appearance of neurological and psychiatric symptoms that may cause suffering to the patient, and even be disabling.

When considering the psychiatric manifestations of infectious illness, it is important to consider clinical manifestations derived

from a possible systemic infection, which can be less obvious than a direct involvement of the central nervous system. Acute organic reactions may accompany many systemic infections, especially at the extremes of life. A clear example is the delirium that frequently occurs with pneumonia in the elderly. In these clinical syndromes, several factors could be responsible for the alterations in cerebral metabolism. The mere fact of having a fever could be involved. Cerebral anoxia often appears to be responsible, or the influence of toxins derived from the infecting micro-organism. More complex metabolic disturbances or the accumulation of toxic intermediate products can also be implicated.

Likewise, infections that course as chronic or subacute illnesses are frequently accompanied by the onset of depressive syndromes. One of the factors implied in clinical depression that occurs within the context of systemic infectious illnesses (e.g. tuberculosis and infectious mononucleosis), is a sense of physical vulnerability, possibly heightened by a loss of strength and negative changes in the patient's appearance. Patients are often afraid of losing their earning capacity or even their jobs, as well as other social and occupational problems associated with the illness.

Another very important factor, above all with the human immunodeficiency virus (**HIV**) and other sexually transmitted disease (**STD**), is the social stigma that these patients may suffer.[1] Sexually transmitted disease infection implies sexual activity that historically carries connotations of illicit, casual, sexual encounters, and acquiring an STD is frequently associated with embarrassment and social stigma.

In addition to the disease itself, the medications commonly used to treat infectious illnesses can have side-effects that alter patients' behaviour, as well as their cognitive and affective functioning (Table 5.3.5.1).

In this chapter we consider infections of clinical interest in the practice of psychiatry. These conditions will be dealt with briefly, and textbooks of general medicine should be consulted for further details. Prion diseases and chronic fatigue syndromes, which are also related to the subject of the present chapter, are discussed in Chapters 4.1.4 and 5.2.7, respectively.

Table 5.3.5.1 Neuropsychiatric adverse effects of drugs frequently used in the treatment of infectious diseases

Drug	Adverse effect
Aciclovir	Headache, somnolence, tremor, confusion, lethargy, seizures, agitation, major depression with psychotic symptoms
Amphotericin B	Delirium
Chloramphenicol	Memory impairment, confusion, depersonalization, hallucinations
Cycloserine	Depression, anxiety, confusion, hallucinations, paranoia, agoraphobia
Didanosine (ddI)	Headache, asthenia, polyneuropathy
Efavirenz	Dizziness, headache, insomnia, inappropriate behaviour, depression, concentration impairment, agitation, abnormal dreaming, and somnolence
Foscarnet	Asthenia
Gentamicin	Confusion, hallucinations
Interferon	Depression, anxiety, irritability, delirium
Isoniazid (INH)	Headaches, vertigo, hyper-reflexia, neuritis, convulsions, ataxia, toxic, encephalopathy, confusion, psychosis, antidepressant effect
Ketoconazole	Somnolence, delirium
Para-aminosalicylate (PAS)	Toxic psychosis
Penicillin G (procaine)	Hallucinations, seizures, agitation, confusion
Rifampicin (rifampin)	Myopathy, headache if hypersensitivity
Streptomycin	Toxic effects on cranial nerve VIII (vestibular), vertigo, nystagmus, ataxia, neuromuscular junction blockade
Sulphonamide	Anxiety, depression, insomnia, hallucinations
Trimethoprim-sulphamethoxazole	Vertigo and confusion
Zalcitabine (ddC)	Polyneuropathy
Zidovudine (AZT)	Headache, myalgia, insomnia, asthenia, somnolence, anxiety, depression, mania, restlessness

HIV infection

Patients infected with HIV are at an increased risk for a variety of mental disorders. Those encountered most frequently in psychiatric practice are discussed below. HIV dementia is discussed in Chapter 4.1.9.

Nature of neuropsychiatric disorders in HIV-infected patients

Neuropsychiatric disorders are common in HIV-infected patients, and they can be either primary or secondary. **Primary** complications are those that can be attributed directly to the infection of the central nervous system by the virus, or to immunopathological events precipitated by HIV infection. Primary HIV-related brain disorders include HIV-related dementia and minor cognitive disorder.[2] Immune suppression can lead to a variety of secondary complications affecting the brain, including opportunistic infections (e.g. cerebral toxoplasmosis and progressive multifocal leucoencephalopathy) and tumours (e.g. cerebral lymphoma). **Secondary** complications in the form of acute and subacute syndromes (e.g. delirium) often occur as a result of cerebrovascular complications and toxic states induced by various therapeutic agents.

HIV-associated acute stress reaction

This transitory syndrome appears in some individuals after they are notified of their seropositivity. It is equally frequent among those who, after a period as an asymptomatic carrier, are informed that the infection has progressed towards full-blown AIDS. The appearance of these symptoms is closely linked in time to the stressful circumstance, and generally remits in hours or days.

The symptoms are highly varied. Some patients suffer from intrusive thoughts or brooding related to their uncertainties regarding health, the future, the risk of contagion to others (especially loved ones), and the idea of death. The vegetative symptoms of panic attacks are also usually present. In more severe cases, the patient may also present social isolation, verbal expressions of rage or feelings of desperation, and other forms of altered behaviour.

Depression

(a) Clinical features

Depression is one of the most common psychiatric disorders found among HIV-infected individuals. Symptomatic stages of HIV infection are associated with an increased prevalence of depressive symptoms and a syndromal diagnosis of major depression.[3]

There are several factors behind the increased morbidity for affective disorders found in this population. First of all, the patient's discovery of the infection has a dramatic **psychological impact**, as does the disease's relentless progression. Second, the **neurotropism of the virus** itself produces neuropathological changes in deep grey structures whose dysfunction is known to cause mood disturbances and changes in the neurotransmission systems, which may contribute to the development of depression. Finally, the groups that in Western countries are at the highest risk for HIV infection (intravenous drug users and male homosexuals/bisexuals) are also known to be at a **high risk for depressive syndromes**, independently of having the virus. The risk factors for depression appear to be similar to those for HIV-seronegative patients and include, besides advanced HIV infection: loss of social support; personal and family history of depression; drug use; and lack of confidants.

When severe physical disease is present the diagnosis of major depression can be difficult to make, because the disease itself may be the real source of many depressive symptoms, for example insomnia, loss of appetite and weight, fatigue, lack of energy, retardation, and concentration difficulties. To avoid misdiagnosing depression, it is important to focus on the more psychological, as opposed to somatic, symptoms associated with low mood. These include **persistent low mood**, **loss of enjoyment** of usually pleasurable activities, **suicidal thoughts** and marked **feelings of hopelessness**, **guilt**, and **self-reproach**. Suicidal ideation may not be expressed directly, but may be expressed more passively, for example poor adherence to medical treatment. Assessment of depressed mood also requires evaluation of the probable contributing factors.

(b) Management

(i) Pharmacological treatment

Antidepressants are the treatment of choice in major depression, as well as in less severe depressive syndromes that are unresponsive to psychological and social intervention. Tricyclic antidepressants have been shown to be effective in treating depressed HIV-positive patients.[4] AIDS patients can respond to lower dosages of tricyclics (25–100 mg), but they may also suffer severe anticholinergic effects at reduced dosages. Therefore, the choice of an antidepressant for these patients should be guided by its side-effect profile.

Several studies have been published showing therapeutic response to selective serotonin reuptake inhibitors in seropositive patients with major depression.[5] Many clinicians prefer the newer drugs in the medically ill, not only because of their higher acceptance among patients, but also because of their greater overdose safety margin.

(ii) Psychotherapy

Psychosocial interventions derived from a wide variety of theoretical orientations are effective in treating depression among individuals infected with HIV. There is good evidence for the value of psychological intervention in the management of HIV patients. Both interpersonal psychotherapy[6] and cognitive–behavioural group therapy[7] may be particularly beneficial for HIV patients with depressive symptoms.

Psychosis

Psychotic disorders sometimes occur in people with HIV infection. While their prevalence is not high, such a development can lead to complicated diagnostic and management problems. The fact that psychosis can be related to HIV infection does not imply that a new disease entity or diagnostic category has been identified. When seropositive individuals present with psychotic symptoms, efforts should be made to clarify the clinical features and to establish their aetiology, which could well be unrelated to HIV. While in some cases the psychotic symptoms may be the result of subtle or gross brain pathology associated with HIV infection, in others it may be iatrogenic or secondary to substance misuse. Psychiatric patients per se may be considered a group at risk for contracting HIV infection.[8]

Neuroleptics are the treatment of choice for controlling psychotic symptoms. The risk of developing antipsychotic-induced extrapyramidal symptoms is higher in psychotic patients with AIDS than in psychotic patients without AIDS. AIDS patients may have an increased risk of developing tardive dyskinesia, neuroleptic malignant syndrome, and severe dystonic reactions.[9] The presence of organic cerebral deterioration, in particular HIV-associated dementia, is a risk factor for the development of neuroleptic malignant syndrome. In general, when using neuroleptics in this population, the best course is to start off with low doses, and increase the dosage slowly and progressively. The new antipsychotic risperidone has been associated with fewer extrapyramidal side-effects and used successfully in this group of patients.[10]

Mania

HIV seems to increase the risk of manic episodes, and mania is a frequent reason for psychiatric hospitalization among people with the virus.[11] In some cases illicit drug use or iatrogenic causes are implicated, for example the chance association of HIV infection and bipolar affective disorders, but generally no obvious aetiological factors can be identified. Mania has been found to be a side-effect of medication frequently used for HIV/AIDS, including **didanosine (ddI)**, **ganciclovir**, **procarbazine**, **estavudine** (d4T), steroids, and **zidovudine** (**AZT**). Most cases of new-onset mania occur in advanced HIV disease and they are often associated with the presence of substantial cognitive impairment. New-onset mania in severe symptomatic disease is predictive of reduced survival.

Standard pharmacotherapy with neuroleptics and lithium are effective, but the usefulness of these drugs may be restricted by the development of severe adverse effects in immunosuppressed HIV-infected patients. Most psychiatrists choose atypical neuroleptics for HIV. However, these agents are not without risk for extrapyramidal side-effects in HIV patients, including the risk for metabolic inhibition of some agents by protease inhibitors. Potent antiretroviral therapy has been documented to protect against the development of HIV-associated mania.[12]

Delirium

Delirium is one of the organic mental disorders observed most frequently in hospitalized HIV-infected patients. The exact prevalence of delirium or acute organic brain syndrome in HIV is unknown.

Table 5.3.5.2 Aetiology of delirium in HIV-infected patients

Infections	Encephalitis due to HIV, syphilis, toxoplasmosis, cryptococcosis, coccidioidomycosis, progressive multifocal leucoencephalopathy, herpesvirus
Abstinence	Alcohol, opiates
Metabolic	Depletion of volume, hydro-electrolytic alterations, transfusions
Hypoxia	Pneumonia with respiratory, compromise
Deficiencies	B-complex vitamins
Cerebral vascular event	
Medication	Anticholinergics, central nervous system depressors
Intracranial mass	Haematoma, neoplasias
Toxic	Drugs of abuse

Patients with advanced systemic disease and dementia are at a high risk for delirium, the cause of which is often multifactorial. The precipitant organic factors involved are listed in Table 5.3.5.2.

A conservative attitude has been recommended for the management of these conditions, with the use of low oral or intramuscular doses of neuroleptics, and correction of the organic disorders responsible for the development of disturbances in the level of consciousness.[13] However, other authors have postulated that patients suffering from delirium and agitation should be given high doses of neuroleptics—alone or in combination with lorazepam—in cases where quick control of the symptoms is vital.[14] The efficacy of pharmacological interventions in patients with delirium is heightened if treatment is begun as soon as the first symptoms appear.

Other central nervous infections in HIV-related illness

In the advanced phases of AIDS, opportunistic infections are highly varied, as are the neoplasias that can develop in immunodepressed individuals, which affect the central nervous system. The more frequents are:

- **Progressive multifocal leucoencephalopathy.** This is a grave neurological complication, linked to papovavirus infection. Dementia can develop rapidly, with focal neurological alterations such as blindness, ataxia, and hemiparesia. Death follows very quickly thereafter, and there is no known treatment. Computerized brain images taken from these patients show a characteristic involvement of the white matter.
- **Cerebral toxoplasmosis.** It is linked to the reactivation of a latent cerebral infection by *Toxoplasma gondii,* an opportunistic intracellular protozoan. The clinical presentation can vary greatly, but it is characterized by the rapid development of a marked alteration in the mental state. The focal involvement can produce headache and lateralized neurological effects. The lesions tend to be located in basal ganglia. Diagnosis is based on structural neuroimaging tests, and treatment is with pyrimethamine and sulphadiazine.
- **Cryptococcal meningitis (torulosis).** This form of meningitis, caused by infestation with the yeast-like fungus *Cryptococcus neoformans,* is characterized by headache, meningism (although it sometimes courses without this symptom), photophobia, nausea, fever, and delirium. The diagnosis is made after a lumbar puncture, and analysis of the culture and antibodies.

Syphilis

A century ago, patients with general paresis due to cerebral syphilitic infections constituted a high proportion of mental hospital admissions, and accounted for an appreciable part of the chronic population of such institutions. With the identification in the early twentieth century of the causative agent, *Treponema pallidum*, and the development of effective methods of treating syphilis, this condition has become relatively rare. Historically, the study of syphilis of the central nervous system has been of great interest to psychiatrists due to the light it sheds on the nature of the relationship between cerebral and mental disease. It was one of the first mental disorders for which a specific organic aetiology was demonstrated, and the first to respond to a medical treatment.

Syphilis remains a major problem in certain areas of the world. Because during its early stages it is a genital ulcerative disease, syphilis facilitates the transmission of HIV and may be particularly important in contributing to HIV transmission in those regions where the rates of both infections are high.

Clinical features

Syphilis is a complex STD with an extremely variable clinical course. Neurosyphilis presents 5 years or more after the initial infection. It affects 10 per cent of non-treated cases, and can take several clinical forms.[15]

- **Asymptomatic neurosyphilis**. Infected subjects have abnormalities in the cerebrospinal fluid (pleocytosis, elevated protein, and reactive VDRL score), but no symptoms or signs of central nervous system disorder. It can evolve into a symptomatic form or remit on its own.
- **Meningovascular syphilis.** Appears within 1 to 5 years of primary infection, although it can occur as early as 6 months and as late as 12 years.[16] In the clinical picture, the patient may develop stroke syndromes of subacute onset with a preceding encephalic picture, including psychiatric disturbances such as lability or personality changes. The patient may complain of headache, lethargy, and malaise, and may experience difficulty in concentration and exhibit faulty judgement. Emotional instability and irritability are common. Mental deterioration may progress to dementia, which can be accompanied by delusional symptomatology and episodes of excitation.[16]
- **General paresis.** This form of parenchymal neurosyphilis is also known as *dementia paralytica* or *general paralysis of the insane.* It usually first appears some 20 years after the initial infection. Its initial symptoms are memory disturbance, dysarthria, and hyper-reflexia, which may be accompanied by personality changes and irritability—in many cases the latter are the presenting abnormalities. The symptoms progress to dementia with abnormal motor function and psychotic symptoms. These organic psychoses were frequent in the pre-penicillin era and are known for their florid clinical picture—prominent euphoric mood, expansive demeanour, and delusions of power, wealth, or social position. Other cases may resemble depressive psychosis with somatic delusions.
- **Tabes dorsalis.** This condition is a degeneration of the ascending fibres from the dorsal root ganglia, resulting in atrophy of the dorsal roots and demyelinization in the posterior columns of the

cord. It can develop from 3 to 20 years after the initial infection. Main symptoms are the loss of position reflexes, ataxia, vibration sense, incontinence, and lacinating pains involving many areas of the body.

Diagnosis and management

The clinical picture of neurosyphilis is so variable that routine serological testing upon admission to psychiatric inpatient units has been recommended. In clinical practice, a rise in atypical syndromes with minor symptomatology has been attributed to partial suppression of the infection during its early stages by antibiotics taken for other reasons.[16] The diagnosis of neurosyphilis should be considered on the basis of the patient's symptoms and clinical signs, and confirmed by serology and analysis of the cerebrospinal fluid.

Penicillin is the drug of choice for *T. pallidum* infections, and in the treatment of neurosyphilis, but the dosage must achieve treponemicidal levels within the cerebrospinal fluid. In untreated cases, death usually occurs within 4 to 5 years. If treatment is given early, the condition usually remits; in already established cases, the progression of the disease can be halted. Antipsychotics are indicated for the symptomatic management of the excitement and psychotic symptoms that these patients may present. Clinicians should have in mind that all patients who have syphilis should be tested for HIV.

Other sexually transmitted diseases

A diverse range of psychological symptomatology is associated with STDs, in which maladaptive or pathological responses to infection (or fear of infection) may occur. Most of the studies that evaluated the psychological effects of having a STD have been carried out in patients attending genitourinary clinics, and focused on genital herpes, a common, recurrent, and painful infection. The response to diagnosis of a STD can include depression, anxiety, anger, social withdrawal, feelings of loneliness, and sexual dysfunction.[17,18] Also, high rates of hypochondriasis and veneroneurosis (a strong but unfounded conviction of having a venereal disease) are found in STD clinics and are frequently associated with psychiatric morbidity. Psychological interventions can effectively reduce the distress associated with STDs, contribute to the control of the infection, increase compliance with medication regimens, and reduce somatic symptoms misattributed to a STD.

Tuberculosis

Clinical picture

In spite of pharmacological advances, tuberculosis continues to be a serious public health problem in many parts of the world, especially due to the increased tuberculosis rate in HIV-infected patients and the appearance of multidrug-resistant tuberculosis.

Tuberculosis patients may present with vegetative signs suggestive of depression, especially the elderly and those in the symptomatic stages of HIV infection. The initial depressive symptoms are weight loss, lethargy, lack of interest, and mental confusion. They may also develop sleep disturbances due to night sweats and nocturnal coughing.

Tuberculosis patients quite frequently present with neuropsychiatric symptoms, which can be related to very different circumstances. First, clinicians should bear in mind that tuberculosis infection often develops in patients who previously had a severe psychiatric pathology, such as alcoholism or intravenous drug abuse, or in the chronic mentally ill. Also, the most common psychiatric symptoms, such as emotional lability and depression, could be related, among other factors, to the feeling of invalidity that accompanies the illness, and its social stigma. In addition, the preventive treatment of those in contact with the patient can trigger feelings of guilt. In some cases, tuberculosis leads to chronic respiratory disease, which is also associated with depressive symptoms, suicidal ideation, and cognitive impairment, particularly in debilitated patients.

Neuropsychiatric disorders in patients with tuberculosis can also be related to cerebral infections: tuberculous meningitis, potentially a very serious complication, develops in 5 per cent of all cases. Since these patients frequently present with mental symptoms that can figure prominently from the outset and even precede overt signs of meningeal infection, the correct diagnosis should be established urgently, in order to institute specific therapy as soon as possible.

Diagnosis and management

Sputum tests, cultures, and chest radiographs, as well as a tuberculin skin test, are standard diagnostic tools. In cases of suspected tuberculous meningitis, a spinal puncture is necessary to determine whether the patient has lymphocytosis and a moderate increase in protein. The diagnosis is confirmed when tubercle bacilli can be identified or cultured from the fluid; however, since irreversible brain damage may result from waiting for cultural confirmation, it is often necessary to begin therapy on the basis of a presumptive clinical diagnosis. The brain scan may show hydrocephalus, focal infarcts, and exudate in basal brain cisterns.[19]

It is important in psychiatric settings that suspected tuberculosis patients receive a proper diagnostic evaluation, not only for the sake of their own health but also for that of other patients who may be exposed to the infection in the unit. It may be necessary to transfer psychiatric inpatients with tuberculosis to a ward where isolation can be assured. The Centres for Disease Control recommend routine tuberculosis screening for patients in HIV risk groups, and for residents of mental health facilities.[20]

The most commonly used drugs for the treatment of tuberculosis are isoniazid, rifampicin, pyrazinamide, ethambutol, and streptomycin. Compliance is vital to achieve effective treatment.

Lyme disease

First described in the United States in 1975, this infection has also been reported in Europe, Australia, and other parts of the world. It is caused by the spirochaete *Borrelia burgdorferi*, which is carried and transmitted by the deer tick. The somatic symptomatology features a characteristic skin lesion, an expanding erythematous annular lesion which usually first appears 3 to 32 days after the initial transmission and may last for several weeks. In 15 per cent of the patients, the disease progresses to a secondary phase marked by neurological symptoms, for example meningoencephalitis, radiculitis, central and peripheral neuropathy, and myelitis.

The neuropsychiatric symptomatology consists in difficulties involving memory, orientation, and calculation. Even years after the first infection, patients can present with violent and impulsive

behaviour, labile affect, and depression. Cases of psychotic or catatonic syndromes and chronic dementia have been described. Many patients with Lyme disease, who suffer from neurological symptoms, present with signs of encephalopathy with alterations in their sleep, affect, and memory. The diagnosis can be established from a serological analysis.

Encephalitis

Encephalitis may be caused primarily by a viral disease affecting the brain or can be a complication of bacterial meningitis, septicaemia, or brain abscesses. It can occur after influenza, herpes simplex, measles, rubella infections, and also after vaccination. In the acute stage, the patient may present with headache, vomiting, and seizures. Patients may develop a confusional syndrome. In rare cases, the encephalitis may present with predominantly psychiatric symptoms. That is the case of herpesvirus encephalitis, which due to its damage to temporal lobes can cause a serious amnestic syndrome.

In clinical practice, the psychiatrist is more likely to see the complications that appear after the acute episode in the form of anxiety and depressive syndromes, personality change, and dementia. In the early years of life, encephalitis may be followed by behavioural disorders.[16]

Infectious mononucleosis

The Epstein–Barr virus causes 90 per cent of all cases of infectious mononucleosis. It can appear at any age, but the illness tends to manifest itself clinically in adolescents and young adults.

The most important symptoms are fever, general malaise, diffuse lymphadenopathies, and laryngitis. However, complications can lead to encephalitis and paralysis of the cranial nerves. In the case of encephalitic compromise, delirium can result. Depressive syndromes have also been observed after an acute infectious episode, accompanied by fatigue.

Brucellosis

This infection is produced by micro-organisms of the genus *Brucella*, and is transmitted by exposure to or ingestion of contaminated animal products, especially unpasteurized milk products, or contact with infected animal tissues. Onset can be insidious, since it mimics other more common illnesses, with low fever, fatigue, and sweating, but 10 to 20 per cent of cases present with splenomegaly. The psychiatric manifestations of the disease can include depressive or anxious syndromes. Diagnosis can be confirmed by blood or lymph cultures or bone marrow biopsy, although the majority of diagnoses are made serologically.

Further information

Fernandez, F. and Ruiz, P. (eds.) (2006). *Psychiatric aspects of HIV/AIDS*. Lippincott Williams & Wilkins, Philadelphia, PA.

APA AIDS resource center: www.psych.org/AIDS

CDC resource center: www.cdc.gov/std

References

1. Forstein, M. (1988). Understanding the psychological impact of AIDS: the other epidemic. *New England Journal of Public Policy*, **4**, 159–73.
2. Catalan, J., Burgess, A., and Klimes, I. (1995). *Psychological medicine of HIV infection*. Oxford Medical Publications, Oxford.
3. Ciesla, J.A. and Roberts, J.E. (2001). Meta-analysis of the relationship between HIV infection and risk for depressive disorders. *The American Journal of Psychiatry*, **158**, 725–30.
4. Markowitz, J.C., Rabkin, J.G., and Perry, S.W. (1994). Treating depression in HIV-positive patients. *AIDS*, **8**, 403–12.
5. Rabkin, J.G., Wagner, G.J., and Rabkin, R. (1999). Fluoxetine treatment for depression in patients with HIV and AIDS: a randomized, placebo-controlled trial. *The American Journal of Psychiatry*, **156**, 101–7.
6. Markowitz, J., Kocsis, J.H., Fishman, B., *et al.* (1998). Treatment of depressive symptoms in human immunodeficiency virus-positive patients. *Archives of General Psychiatry*, **55**, 452–7.
7. Blanch, J., Rousaud, A., Hautzinger, M., *et al.* (2002). Assessment of the efficacy of a cognitive-behavioral group psychotherapy programme for HIV infected patients referred to a consultation-liaison psychiatry department. *Psychotherapy and Psychosomatics*, **71**, 77–84.
8. Ayuso-Mateos, J.L., Montañes, F., Lastra, I., *et al.* (1997). HIV infection in psychiatric inpatients: an unlinked anonymous study. *British Journal of Psychiatry*, **170**, 181–5.
9. Ayuso, J.L. (1994). Use of psychotropic drugs in patients with HIV infection. *Drugs*, **4**, 599–610.
10. Singh, A.N., Golledge, H., and Catalan, J. (1997). Treatment of HIV related psychotic disorders with risperidone: a series of 21 cases. *Journal of Psychosomatic Research*, **42**, 489–93.
11. Kierburz, K., Zettelmaier, A., Ketonen, L., *et al.* (1991). Manic syndrome in AIDS. *The American Journal of Psychiatry*, **148**, 1068–70.
12. Mijch, A.M., Judd, F.K., Lyketsos, C.G., *et al.* (1999). Secondary mania in patients with HIV infection: are antiretrovirals protective? *Journal of Neuropsychiatry and Clinical Neurosciences*, **11**, 475–80.
13. Breitbart, W., Marotta, R., Platt, M.M., *et al.* (1996). A double-blind trial of haloperidol, chlorpromazine and lorazepam in the treatment of delirium in hospitalised AIDS patients. *The American Journal of Psychiatry*, **153**, 231–7.
14. Fernandez, F., Levy, J.K., and Mansell, P.W. (1989). Management of delirium in terminally ill AIDS patients. *International Journal Psychiatry in Medicine*, **19**, 165–72.
15. Simon, R.P. (1985). Neurosyphilis. *Archives of Neurology*, **42**, 606–13.
16. Lishman, W.A. (1998). *Organic psychiatry* (3rd edn). Blackwell Science, Oxford.
17. Catalan, J., Bradley, M., Gallwey, J., *et al.* (1981). Sexual dysfunction and psychiatric morbidity in patients attending a clinic for sexually transmitted diseases. *British Journal of Psychiatry*, **138**, 292–6.
18. Hedge, B. (1997). Sexually transmitted diseases. In *Cambridge handbook of psychology, health and medicine* (eds. A. Baum, S. Newman, J. Weinman, R. West, and C. McManus), pp. 584–5. Cambridge University Press, Cambridge.
19. Rovira, M., Romero, F., and Torrent, O. (1980). Study of tuberculosis meningitis by CT. *Neuroradiology*, **19**, 137–41.
20. Dooley, S.W., Castro, K.G., Hutton, M.D., *et al.* (1990). Guidelines for preventing the transmission of tuberculosis in health-care settings, with special focus on HIV-related issues. *Morbidity and Mortality Weekly Report*, **39** (RR-17), 1–29.

5.3.6 Psychiatric aspects of surgery (including transplantation)

S. A. Hales, S. E. Abbey, and G. M. Rodin

Attention to psychiatric disturbances and to emotional distress is important in the surgical setting, from the time of the initial diagnostic assessment, to the perioperative period and the phase of subsequent recovery and rehabilitation. Psychiatric illness and psychological factors, which are not taken into account prior to surgery, may contribute to inaccurate diagnoses, unrealistic assessment of the surgical risk, unnecessary surgery, and complications that could have been avoided or minimized. This chapter will address these factors and provide an approach to the consideration of psychiatric factors and interventions in this setting.

Preoperative assessment and intervention

The assessment of all patients being considered for surgery should include a brief evaluation of their current emotional state, cognitive functioning, personal circumstances, present or past history of psychiatric illness, and personality and coping style, as these factors may affect their adjustment to surgery. Psychiatric consultation may be indicated for a number of specific reasons discussed below.

Psychological contributors to the patient's physical symptoms

The most common psychological factor that complicates the surgical assessment is a low pain threshold and a tendency to somatize, i.e. to experience and communicate emotional distress in physical terms. When emotional factors amplify somatic symptoms, it is more difficult to distinguish organic from functional disorders on the basis of the clinical history. At the extreme end of the continuum of somatization is the dramatic presentation of physical symptoms, which may mimic a surgical condition, in the absence of organic disease. This syndrome may fulfil criteria for a somatoform disorder, such as a somatization disorder, conversion disorder, or somatoform pain disorder.[1] In such cases, careful attention to the objective indications for surgery is required. Individuals with a body dysmorphic disorder, a syndrome of perceived or imagined ugliness, may present with repeated requests for cosmetic surgery.[2] However, cosmetic surgery is unlikely to relieve the body dissatisfaction of such patients, whose condition has much in common with obsessive–compulsive disorder. Dissatisfaction with the results of surgery is common in patients with body dysmorphic disorder, and litigiousness and threats towards the treating surgeon may occur in a small proportion of cases. In general, the failure of clinicians to consider the contribution of somatization to the clinical presentation may lead to unnecessary or inappropriate surgery. When this occurs, postsurgical complications may interfere with the subsequent evaluation of persistent physical symptoms. Somatoform disorders are discussed further in Chapter 5.2.2.

Although physicians commonly consider that emotional factors may amplify physical symptoms, the possibility that disease has been intentionally simulated or fabricated is usually not entertained. Factitious disorder refers to a syndrome in which such behaviour is enacted for no apparent reason, other than to assume the patient role. This is in contrast with malingering, in which there may be reports of physical symptoms motivated by the desire for some specific secondary gain, which may be financial or compensation-related. While relatively rare, factitious disorder poses particular challenges, in both detection and treatment, to psychiatric and surgical teams. Individuals with this disorder may produce or simulate disease in various ways, such as by self-inflicting wounds that require surgical intervention or by surreptitiously contaminating themselves to produce infection. Patients with this condition may also communicate plausible symptoms of a surgical condition. Such patients are at risk to receive unnecessary surgery and should be regarded as suffering from a serious and potentially life-threatening condition. The majority of such patients are unwilling to accept psychological or psychiatric assistance, but an ongoing, supportive relationship with a medical caregiver may diminish this symptom pattern. Factitious disorder is discussed further in Chapter 5.2.9.

Capacity to consent to surgery

Providing information and obtaining informed consent are routine and essential aspects of preoperative care provided by the surgical team. Informed consent to a surgical procedure requires disclosure of pertinent information by the treating physician, and understanding of the information, decisional capacity, and voluntary choice on the part of the patient.[3] The decisional capacity of patients depends on their ability not only to understand information relevant to the decision, but also to apply it to their own situation and to express a consistent voluntary choice.[4] Information required by patients to make an informed surgical decision includes the rationale, risks, and benefits of the surgery, the potential alternative treatments, and the risk of not proceeding with surgery. In most jurisdictions, the emergency treatment of incapable persons is permitted, when substitute consent is not available, unless the clinician has reason to believe that the person would refuse such treatment if he or she were capable.[3] When it is not an emergency, substitute consent must be obtained on behalf of individuals who are incapable of providing informed consent. The legal requirements for substitute consent vary in different jurisdictions.

The capacity to provide informed consent may be impaired by cognitive dysfunction, by psychiatric illness, or by contextual factors, such as the clarity and relevance of the information disclosed or the manner of disclosure. If screening by the treating surgeon indicates that the patient may be incapable, a psychiatric consultation may be requested to evaluate the patient's decisional capacity.[5] Patients with cognitive impairment or a major psychiatric disorder, such as schizophrenia, are not necessarily incapable of making treatment decisions, unless these conditions affect their understanding and appreciation of information relevant to the decision. Numerous tools have been developed to assess decisional capacity but there is no current gold standard.[6]

Obtaining informed consent for surgery is not only a legal and ethical requirement, but also a crucial dimension of the surgeon–patient relationship. In some cases, treatment refusal reflects a breakdown in the relationship between the surgeon and the patient more than it does an informed decision of the patient to reject a

recommendation for surgery. When this occurs, attention to the physician–patient relationship, and the provision of additional information, may help to relieve the impasse so that an informed decision can be made. In other cases, treatment of a major psychiatric illness, such as a psychotic episode in a patient with schizophrenia, is necessary to restore the patient's capacity to provide consent.

Assessment of the response to surgery

Surgical patients face numerous stressors, including the fear of pain, disfigurement, and the loss of control, as well as the possibility of major medical complications and death. The response to these stressors may be affected by the nature of the illness and the surgical procedure, its personal meaning, the prior history of trauma, the support which is anticipated and perceived from medical caregivers and significant others, and the prior experience of the individual with medical or surgical procedures. The age and life stage of the individual, the risk associated with the procedure, and the prognosis of the underlying or associated medical conditions may also affect the psychological response in the perioperative period. Apprehension and mistrust are more common in those who have previously suffered from the adverse effects of missed or delayed diagnosis or treatment. Attitudinal factors, including positive expectations and the desire to participate actively in the recovery process, may also affect clinical outcomes. The desire to maintain a sense of control may be adaptive during the preparation and rehabilitation phases but may be associated with greater distress immediately following surgery, when there is an inescapable and predominant requirement to depend on others.[7] Those with more attachment anxiety, i.e. concern about the availability of support from others, may benefit from predictability and reliability in relation to caregivers, whereas those who tend to be more self-sufficient may benefit most from strategies which promote self-reliance and self-care.[8, 9]

There has been particular interest in psychosocial issues in the setting of transplantation surgery. This occurred, in part, because of the desire of transplant programmes to select optimal candidates for organ transplants, which were experimental and/or in scarce supply. However, the psychiatric and psychosocial selection criteria for transplant surgery have become less stringent, as the transplantation of particular organs has become more routine.[10] At present, psychosocial evaluation of transplant candidates by a multidisciplinary team allows for the identification, treatment, and monitoring of factors that may affect compliance, morbidity, and psychosocial outcomes. Organ transplants from living donors, such as for bone marrow, kidney, and liver transplantation, are unlike most other surgical interventions in that they necessitate surgery for individuals without pre-existing disease. Psychosocial evaluation of such donors includes consideration of the process of decision-making and informed consent, the adaptive capacities of the individual, the degree of social support, and the relationship of the donor to the recipient.[11] Although there has been concern about the psychological consequences of such surgery, the available evidence suggests that organ donation is usually well tolerated and experienced in positive terms by the donor, particularly when the surgical and medical outcomes are favourable.[12]

There is now increasing evidence that the systematic preoperative education of patients and their family caregivers in a therapeutic context may enhance adjustment to surgical procedures.[13] Postoperative education may also improve subsequent rehabilitation following surgical procedures. Such approaches are consistent with modern Western trends towards consumerism and patient empowerment, in which greater emphasis is placed on assisting patients to assume more responsibility for their medical course and treatment outcome. This approach has also been necessitated by the trend towards earlier hospital discharge of surgical patients into the community where much more self-care is required.

Anxiety

Preoperative anxiety is common and may be particularly problematic in patients awaiting procedures such as transplantation, which usually occur in the course of a life-threatening condition, and are associated with long and unpredictable waiting periods for surgery. Anxiety has been reported to be more common in younger patients, in females, and in those who are unmarried or who have less perceived social support.[14] Research suggests that preoperative anxiety may complicate postoperative recovery through behavioural and physiological mechanisms.[15] Symptoms of anxiety can usually be managed with education and reassurance, but when they are persistent and problematic, interventions such as progressive relaxation and guided imagery may be helpful both to reduce symptoms of anxiety and to enhance feelings of self-control.[16] Some patients benefit from a benzodiazepine to reduce preoperative anxiety, but those with antecedent anxiety disorders may require more intensive intervention, as outlined in Chapter 4.6.1. Prior to elective procedures, patients with specific blood or needle phobias may benefit from systematic desensitization. Those with comorbid panic disorder may require a higher dose of anxiolytic medication in the preoperative period. The surgical team should be aware of such treatment so that the medication can be restarted promptly after surgery, to avoid symptoms of withdrawal and anxiety. If oral medications cannot be reinstituted for a prolonged period of time after surgery, intramuscular lorazepam or intravenous lorazepam or diazepam may be used.

Mood disorders

It is important to detect and treat mood disorders prior to surgery because they are associated with increased surgical morbidity and mortality, and with reduced treatment compliance in the postoperative period.[17] Anaesthetists must be aware of any drugs taken to treat and prevent bipolar and depressive disorders, because some can significantly prolong muscle paralysis secondary to neuromuscular blockade. Furthermore, attention should be paid to serum lithium levels and signs of lithium toxicity since they may be affected by the patient's fluid and volume status. Conventional heterocyclic antidepressants and selective serotonin reuptake inhibitors can be continued until the time of surgery, and then restarted postoperatively, when oral medications can be tolerated. Selective serotonin reuptake inhibitors are known to affect platelet serotonin levels and platelet aggregation and to be occasionally associated with prolonged bleeding times, increased perioperative blood loss, and an increased subsequent need for transfusion.[18] Patients receiving monoamine oxidase inhibitors (MAOIs) are usually advised to discontinue this medication for 1 to 2 weeks prior to their surgery, although this recommendation must be weighed against the risk of withdrawal symptoms and of precipitating

a current depression. The medical charts of patients receiving MAOIs should be clearly labelled to advise that all drugs administered should be screened for their interactions with these drugs and that pethidine (meperidine) and dextromethorphan in particular should not be prescribed due to risk of a serotonin syndrome, which is associated with gastrointestinal, neurological, cardiovascular, and psychiatric symptoms. In addition, hospital charts of patients taking MAOIs should be clearly marked to indicate that they must avoid foods containing tyramine. Patients with bipolar disorder should be monitored for mood alterations since they may be at risk of developing hypomania or mania when steroid medications are used following transplantation surgery.

Psychotic disorders

These disorders, most commonly related to schizophrenia or bipolar disorder, pose challenges which vary depending upon the requirements of the surgery and the patient's mental status. Patients with schizophrenia may be at increased perioperative risk for hypotension, hypothermia, confusion, infection, and for ileus, in those who undergo abdominal surgery. These complications may occur due both to pathology of the endocrine, immune, and cardiovascular systems associated with schizophrenia and to the effects of antipsychotic medications.[19] Those who are being treated with a low-potency antipsychotic may be switched to a high-potency agent to decrease the risk of hypotension, particularly with cardiovascular surgery in the postoperative period. For their own comfort and for that of others, individuals who are actively psychotic may require special arrangements, such as a single room, close or constant observation, and, when feasible, greater family involvement. Such patients require closer monitoring for many reasons, including the increased likelihood that they may misinterpret common ward events as threatening and because they may be at increased risk for exacerbation of their underlying condition and for the occurrence of delirium.

Cognitive disorders

The capacity of patients to understand information and to provide a coherent account of their symptoms is fundamental to the process of diagnosis, informed consent, and assessment of the indications for surgery and the risk of specific postoperative complications. Cognitive impairment prior to surgery may complicate these processes and may be associated with an increased risk of delirium or dementia in the postoperative period. In such cases, neuropsychological testing may be indicated prior to elective surgery to establish a baseline, to assist in the evaluation of decisional capacity, and, to aid in the prediction of postoperative delirium or worsening of dementia.

Personality disorders

Patients with personality disorders are more likely to have greater difficulty than others adapting to the multiple and unpredictable stresses associated with surgery. Those with impulsivity may have difficulty adhering to the preoperative and postoperative regimen and those who are suspicious and mistrustful may be more limited in their ability to form effective treatment relationships and to make treatment decisions, in which an enormous degree of trust is required. Those with a borderline personality disorder may be highly sensitive to feelings of personal injury or neglect and may tend to idealize some caregivers and to denigrate others. These responses may create problematic divisions amongst the treatment team and may adversely affect the care of the patient. A psychiatric consultant may be of help to provide patient support and to educate staff about the underlying psychiatric disorder in such individuals, who may otherwise be viewed negatively by staff who regard their behaviour as simply wilful or manipulative.

Substance abuse disorders

The preoperative assessment should include enquiry about substance use, in order to adjust current medication appropriately and to prevent the occurrence of postoperative withdrawal syndromes. Consultation psychiatrists may play a role in continuing medical education about the value of routine, non-judgemental screening for substance use in medical patients, and about the importance of recording these details in the medical record.

Assessment of psychotropic medications

The pharmacological effects of psychotropic medications must be taken into account in the perioperative period. Important factors that affect the risks and benefits associated with psychotropic medication use perioperatively include:

1 End-organ sensitivity to side effects based on medical comorbidity and organ dysfunction;

2 direct effects of psychotropic medications and their potential interactions with anaesthetic and analgesic agents likely to be prescribed;

3 route of access available (oral, suppository, subcutaneous, intramuscular, intravenous);

4 risk of withdrawal symptoms and recurrence or relapse of a psychiatric disorder if psychotropic medications are to be discontinued.[20]

Information on drug–drug interactions is constantly being updated and current information can be obtained from internet sites such as http://www.drugdigest.org/DD/Home or http://search.medscape.com/drug-reference-search.

Postoperative complications and interventions

Agitation and delirium

Agitation is a common postoperative problem, the frequency of which depends upon the characteristics of the disease, the nature of the surgery and its complications, and the pre-existing vulnerability of the patient. A concerted effort should be made to identify the source of the agitation, which may be a worsening of the medical condition, inadequately controlled pain, or delirium. Delirium, which is described in more detail in Chapter 4.1.1, is a common complication, which develops in more than one-third of cases following surgery.[21] Higher rates of delirium are found following longer procedures, due to intraoperative hypoxemia, following cardiac surgeries, due to hypoperfusion and microemboli formation, following orthopaedic procedures, due to fat emboli, and following cataract surgery, due to the impact of vision loss and of ophthalmic drugs with anticholinergic side effects.[22]

Measures to prevent and ameliorate delirium include the identification and treatment of predisposing risk factors and reversible causes, symptomatic treatment, and environmental interventions to reduce distress and agitation. The latter may include measures to prevent sensory deprivation and disorientation, to monitor safety, and to educate patients and family members about the condition. When distress or agitation associated with delirium threaten the safety and care of the patient, pharmacologic interventions may be necessary. There have been trials of cholinesterase inhibitors[23] and of atypical antipsychotics in those patients who can take oral medications, but haloperidol remains the first-choice medication for management of delirium-associated agitation. It is preferred because it has fewer active metabolites and fewer anticholinergic and sedative effects than other antipsychotic medication and because it can be administered intravenously. This route of administration is usually safe, although arrhythmias with its intravenous use have been reported in patients with histories of alcohol abuse or with cardiomyopathy.[24] Benzodiazepines should be used to treat withdrawal from alcohol and sedative-hypnotics using standard protocols.

Delirium may be highly distressing for both patients and family caregivers.[25] Some patients subsequently retain disturbing memories of their experience during a delirium. Such traumatic recall may also occur, in rare instances, when there has been inadequate anaesthesia. In these cases, patients and their family caregivers typically appreciate a discussion of their concerns and the opportunity to review these events with the surgeon and the anaesthetist. A small number of patients with these disturbing experiences develop symptoms of post-traumatic stress disorder and may benefit from a brief course of psychotherapy or pharmacotherapy to alleviate their symptoms.

Ventilator weaning

Some patients experience difficulty being weaned from the ventilator due to anxiety, depression, delirium, or other psychological factors related to their disease or to the ICU environment. Behavioural approaches to facilitate weaning may include relaxation techniques that do not depend on observation or manipulation of breathing, guided imagery, and/or biofeedback.[26] Weaning problems due to anxiety typically respond to benzodiazepines or to haloperidol, administered in a single dose prior to weaning. Apathetic or depressed patients who have difficulty being weaned may benefit from a psychostimulant.[27] When delirium is the cause of weaning problems, its underlying cause should be identified and treated.

Pain management

Effective pain management is a fundamental aspect of postoperative treatment and may reduce distress, agitation, sleep disturbance, anxiety, mood symptoms, and behavioural disorders. Suboptimal pain management may occur due to inadequate assessment of this symptom, insufficient knowledge of the pharmacokinetic and pharmacodynamic properties of analgesic medication, and unfounded concerns about 'addiction'. Further, some patients refuse to accept adequate analgesia because of misconceptions or personal beliefs regarding the importance of stoicism, vigilance, or personal control. Consultant psychiatrists may help patients to address their concerns about analgesic medication and analgesic adjuvants and may act also as advocates or intermediaries for patients to ensure adequate analgesia.

Sleep disturbance

Sleep disturbances, including reduced total sleep time, fragmentation of sleep, frequent arousals and awakenings, and reduced slow-wave sleep, are common in the immediate postoperative period. These disturbances may be caused by multiple factors, including the noise, temperature and light of the hospital environment, neuroimmunological and other changes associated with the surgical insult, and anaesthetic and analgesic medications.[28] Daytime sleeping, related to prolonged bed rest, lack of intellectual and social stimulation, and reduced circadian cues, may disrupt normal sleep chronobiology and increase difficulty with night-time sleep.

Treatment of sleep disturbances should be directed to the identified cause. In those patients in whom anxiety or the intrusiveness of the hospital environment is the main cause of the sleep disturbance, benzodiazepines may be temporarily used. When there are concerns about substance abuse, newer non-benzodiazepine hypnotics may be preferable, due to the lower risk of tolerance associated with their use.[28] Caution should be exercised with patients for whom benzodiazepine-induced nocturnal respiratory compromise may be problematic or with elderly patients who are susceptible to cognitive compromise or to falls.

Cognitive impairment

Cognitive impairment is a common short-term and long-term complication of major surgery, particularly in those with more advanced age. Gradual recovery of cognitive functions occurs in most patients within 3 months after surgery, although it may take as long as 6 to 12 months.[29] Delirium that is slow to resolve may be an early sign of an associated dementing illness or of a cerebral insult that has occurred during or after surgery. Cognitive impairment may be permanent when there has been irreversible brain damage due to neurosurgery or perioperative complications, such as hypoxia or stroke. Neuropsychological testing may be helpful to document or track changes in cognitive functioning. Education of primary caregivers about the risk and manifestations of cognitive impairment is essential, and institution of home supports and respite care for patients and families facing this complication may be necessary. Some patients with cognitive impairment may also benefit from neurorehabilitation interventions.

Adjustment issues

Longer-term problems in adjustment may occur, particularly following disfiguring surgeries, such as facial surgery, amputations, ostomies, or following procedures such as organ transplantation, which impose complicated postoperative regimens. Sexual difficulties that result from procedures that compromise the neural input or functional integrity of genital structures, or that negatively affect body image or feelings of attractiveness, may be disturbing to patients and spouses who may benefit from specific enquiry and assistance. Patients and their families should be informed about the possibility of adjustment problems with such surgeries and should be given information about available resources. A minority of patients develop clinically significant mood or anxiety disorders or significant compliance problems during the

rehabilitation phase which may necessitate psychiatric consultation and intervention.

Prolonged dysfunction

Prolonged and disproportionate pain and disability occur in a subset of patients. In some of these cases, the surgery was undertaken to relieve refractory symptoms that subsequently proved to be functional or medically unexplained. This may occur with hysterectomies performed to relieve pelvic pain and in surgery to relieve chronic back pain. In other cases, persistent disproportionate symptoms may be perpetuated by secondary gain of a financial or social nature, or by opiate dependence. Undiagnosed and untreated mood, anxiety, and somatoform disorders may also contribute to persistent symptoms. The consultation psychiatrist may be called upon to identify these factors and to help distinguish them from undetected medical/surgical pathology or from the effects of inadequate pain regimens.

Further information

Levenson, J.L. (ed.) (2005). *American psychiatry publishing textbook of psychosomatic medicine*. American Psychiatric Press, Washington, DC.

Lloyd, G. and Guthrie, E. (eds.) (2007). *Handbook of liaison psychiatry*. Cambridge University Press, Cambridge.

References

1. Abbey, S.E. (2002). Somatization and somatoform disorders. In: *The American psychiatric publishing textbook of consultation-liaison psychiatry* (2nd edn) (ed. M.G. Wise and J.R. Rundell), pp. 361–92. American Psychiatric Press, Washington, DC.
2. Phillips, K.A., McElroy, S.L., Keck, P.E. Jr., *et al.* (1993). Body dysmorphic disorder: 30 cases of imagined ugliness. *The American Journal of Psychiatry*, **150**, 302–8.
3. Etchells, E., Sharpe, G., Walsh, P., *et al.* (1996). Bioethics for clinicians: 1. consent. *CMAJ: Canadian Medical Association Journal*, **155**, 177–80.
4. Grisso, L.B. and Appelbaum, P.S. (1998). *Assessing competence to consent to treatment: a guide for physicians and other health professionals.* Oxford University Press, New York.
5. Etchells, E., Sharpe, G., Elliott, C., *et al.* (1996). Bioethics for clinicians: 3. Capacity. *CMAJ: Canadian Medical Association Journal*, **155**, 657–61.
6. Dunn, L.B., Nowrangi, M.A., Palmer, B.W., *et al.* (2006). Assessing decisional capacity for clinical research or treatment: a review of instruments. *The American Journal of Psychiatry*, **163**, 1323–34.
7. Rosenberger, P.H., Jokl, P., and Ickovics, J. (2006). Psychosocial factors and surgical outcomes: an evidence-based literature review. *The Journal of the American Academy of Orthopaedic Surgeons*, **14**, 397–405.
8. Hunter, J.J. and Maunder, R.G. (2002). Using attachment theory to understand illness behavior. *General Hospital Psychiatry*, **23**, 177–82.
9. Tan, A., Zimmermann, C., and Rodin, G. (2005). Interpersonal processes in palliative care: an attachment perspective on the patient–clinician relationship. *Palliative Medicine*, **19**, 143–50.
10. Olbrisch, M.E., Benedict, S.M., Ashe, K., *et al.* (2002). Psychological assessment and care of organ transplant patients. *Journal of Consulting and Clinical Psychology*, **3**, 771–83.
11. Olbrisch, M.E., Benedict, S.M., Haller, D.L., *et al.* (2001). Psychosocial assessment of living organ donors: clinical and ethical considerations. *Progress in Transplantation*, **11**, 40–9.
12. Shrestha, R. (2003). Psychosocial assessment of adult living donors. *Liver Transplantation*, **9**, S8–11.
13. Walker, J. (2002). Emotional and psychological preoperative preparation in adults. *British Journal of Nursing*, **11**, 567–75.
14. Karanci, A.N. and Dirik, G. (2003). Predictors of pre- and postoperative anxiety in emergency surgery patients. *Journal of Psychosomatic Research*, **55**, 363–9.
15. Kiecolt-Glaser, J.K., Page, G.G., Marucha, P.T., *et al.* (1998). Psychological influences on surgical recovery: perspectives from psychoneuroimmunology. *The American Psychologist*, **53**, 1209–18.
16. Horne, D. (1994). Preparing patients for invasive medical and surgical procedures. 2: using psychological interventions with adults and children. *Behavioral Medicine*, **20**, 15–21.
17. Rodin, G., Craven, J., and Littlefield, C. (1991). *Depression in the medically ill: an integrated approach*. Brunner–Mazel, New York.
18. Movig, K.L., Janssen, M.W.H.E., de Waal Malefijt, J., *et al.* (2003). Relationship of serotonergic antidepressants and need for blood transfusion in orthopedic surgical patients. *Archives of Internal Medicine*, **163**, 2354–8.
19. Kudoh, A. (2005). Perioperative management for chronic schizophrenic patients. *Anesthesia and Analgesia*, **101**,1867–72.
20. Huyse, F.J., Tuow, D.J., van Schijndel, R.S., *et al.* (2006). Psychotropic drugs and the perioperative period: a proposal for a guideline in elective surgery. *Psychosomatics*, **47**, 8–22.
21. Dyer, C.B., Ashton, C.M., and Teasdale, T.A. (1995). Postoperative delirium. A review of 80 primary data-collection studies. *Archives of Internal Medicine*, **155**, 461–5.
22. Winawer, N. (2001). Postoperative delirium. *The Medical Clinics of North America*, **85**, 1229–39.
23. Overshott, R., Burns, A., and Karim, S. (2005). Cholinesterase inhibitors for delirium. (Protocol) Cochrane Database of Syst Rev [serial online] [cited 2005 Apr 20] Issue 2. Art. No.: CD005317. DOI: 10.1002/14651858.CD005317. Available from: URL:http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD005317/frame.html
24. Schwartz, T. and Masand, P. (2002). The role of atypical antipsychotics in the treatment of delirium. *Psychosomatics*, **43**, 171–4.
25. Breitbart, W., Gibson, C., and Tremblay, A. (2001). The delirium experience: delirium recall and delirium-related distress in hospitalized patients with cancer, their spouses/caregivers, and their nurses. *Psychosomatics*, **43**, 183–94.
26. Hannich, H.J., Hartmann, U., Lehmann, C., *et al.* (2004). Biofeedback as a supportive method in weaning long-term ventilated critically ill patients. *Medical Hypotheses*, **63**, 21–5.
27. Johnson, C.J., Auger, W.R., Fedullo, P.F., *et al.* (1995). Methylphenidate in the 'hard to wean' patient. *Journal of Psychosomatic Research*, **39**, 63–8.
28. Morin, A.K., Jarvis, C.I., and Lynch, A.M. (2007). Therapeutic options for sleep-maintenance and sleep-onset insomnia. *Pharmacotherapy*, **27**, 89–110.
29. Dijkstra, J.B. and Jolles, J. (2002). Postoperative cognitive dysfunction versus complaints: a discrepancy in long-term findings. *Neuropsychology Review*, **12**, 1–14.

5.3.7 Psychiatric aspects of cancer

Jimmie C. Holland and Jessica Stiles

Introduction

Psycho-oncology addresses the two major psychiatric and psychological dimensions of cancer: first, the responses of patients and their families at all stages of disease and the psychological stresses on health professionals delivering their care. The patient and physician relationship, dependent on effective communication, impacts

the care of all patients, at every visit, at all sites and stages of cancer, and during all treatments. The second dimension addresses the psychological, behavioural, and social factors that influence cancer risk, detection, and survival.

Many cancer centres and hospitals now have multi-disciplinary psychosocial teams consisting of clinicians and clinical investigators from psychology, psychiatry, social work, nursing, and clergy. These teams provide consultation for patients and their caregivers, psychosocial education for oncology staff, and collaboration in studies in which quality of life is important. In addition, active research in brain, immune, and endocrine links is occurring, particularly in the mechanism of cytokines in producing 'sickness behaviour' that may provide a biological basis for common symptoms of fatigue, depression, anxiety, weakness, and cognitive chances in cancer patients.[1, 2]

Despite the fact that many cancer centres and oncology divisions now have a psycho-oncology or psychosocial unit, only a few centres have programmes that include both research and training.

This chapter describes the common psychiatric disorders and psychosocial challenges experienced by cancer patients and the range of interventions available.

Psychiatric disorders

A key challenge for the oncologist is the differentiation of expected, tolerable, transient distress associated with cancer, such as fear, worry, and sadness, from excessive, disabling, persistent distress requiring therapeutic intervention. Most psychiatric disturbances in patients with cancer relate to their illness or treatment side effects.[3] One-third of patients will experience distress that requires evaluation and treatment.[3–6] The percentage is greater among younger patients, those with sites of cancer with poorer prognosis, for example, brain, pancreas, lung, and those who are hospitalized with greater level of illness causing confusional states and greater anxiety and depression.[7,8]

Anxiety

Anxiety is the most common form of distress experienced by patients in the oncology setting (Table 5.3.7.1). It occurs with abnormal metabolic states: hypoxia, pulmonary embolus, sepsis, delirium, bleeding, cardiac arrhythmia, and hypoglycemia. Hormone-secreting neoplasms that produce psychiatric symptoms consistent with mood or anxiety disorder are pheochromocytoma, thyroid tumour, carcinoid, parathyroid adenoma, adrenocorticotropic hormone-producing tumour, insulinoma, and paraneoplastic syndrome, an immunologic non-metastatic central nervous system complications of several tumours (particularly, lung and ovary) that may present with mood or cognitive changes.

Numerous medications produce symptoms of anxiety: corticosteroids, neuroleptics, bronchodilators, thyroxine, and psychostimulants. The antiemetics, including metoclopramide and prochlorperazine, which are widely used for chemotherapy-related nausea and vomiting, produce restlessness, akathisias, and dystonias. Benzodiazepines promptly reduce the restless movements, anxiety, and agitation. Withdrawal states from alcohol, benzodiazepines, sedative-hypnotics, and opioids produce anxiety as prominent symptoms.

Some patients undergoing cyclic chemotherapy receiving highly emetogenic regimens develop anticipatory anxiety, nausea, and vomiting days to hours in advance of receiving the next cycle of treatment.[9–11] More effective antiemetic regimens have significantly reduced the frequency and severity of this problem. However, behavioural interventions paired with antianxiety medications continue to assist in providing relief from this distress.

Table 5.3.7.1 Causes of anxiety in patients with cancer

Situational
Diagnosis of cancer, prognosis discussion
Crisis, illness/treatment
Conflicts with family or staff
Anticipating a frightening procedure
Awaiting results of tests
Fears of recurrence after completing treatment
Disease-related
Poorly controlled pain
Abnormal metabolic states
Hormone secreting tumors
Paraneoplastic syndromes (remote CNS effects)
Treatment-related
Frightening or painful procedures (MRI, scans, wound debridement)
Anxiety-producing drugs (antiemetic neuroleptics, bronchodilators)
Withdrawal states (opioids, benzodiazepines, alcohol)
Conditioned (anticipatory) anxiety, nausea, and vomiting with cyclic chemotherapy
Exacerbation of preexisting anxiety disorder
Phobias (needles, claustrophobia)
Panic or generalized anxiety disorder
Posttraumatic stress disorder (Holocaust survivors, Vietnam veterans, recall of the death of a relative with cancer)
Obsessive compulsive disorder

Patients who have pre-existing phobias, panic attacks, generalized anxiety disorder, or obsessive–compulsive disorder are at risk of experiencing symptom exacerbations during treatment (Table 5.3.7.1).[12] Phobias of needles, blood, hospitals, magnetic resonance imaging machines, or radiation simulators complicate a patient's ability to tolerate hospital procedures or adhere to recommended treatments. Panic attacks superimposed on physical symptoms of dyspnea and tachycardia may be partially alarming to patients.[3,13] Patients with previous traumatic experiences may suffer a recurrence of intrusive re-experiences of painful memories, maladaptive avoidant behaviour or withdrawal, and hypervigilance.[14,15]

Cancer patients with OCD may have increased difficulty during treatment. Intrusive fears may lead to indecisiveness regarding treatment options and reluctance to accept interventions with known therapeutic efficacy. Excessively time-consuming rituals may interfere with a patient's adherence to medical appointments. Inflexibility of thought, hostility, overwhelming distress, and occasionally poor insight contribute to the challenge of engaging these patients and assisting them in accepting interventions.

Management. Anticipatory anxiety prior to medical interventions responds to empathic validation of the fear, adequate preparation to set realistic expectations for the encounter, and rehearsal of the dreaded event.

Significant disabling anxiety symptoms are frequently treated pharmacologically with benzodiazepines, selective serotonin-reuptake inhibitors (SSRIs), mirtazapine, venlafaxine, buspirone, antihistamines, beta-blockers, or neuroleptics. Table 5.3.7.2 outlines the benzodiazepines commonly used and their initial and

Table 5.3.7.2 Common anxiolytic agents

Drug	Brand name	Starting dose/day	Theraputic dose/day
SSRIs			
Sertraline	Zoloft	25–50 mg AM	50–150 mg
Fluoxetine	Prozac	10–20 mg AM	20–60 mg
Paroxetine	Paxil	10–20 mg	20–60 mg
Citalopram	Celexa	10–20 mg	20–60 mg
Escitalopram	Lexapro	5–10 mg	10–30 mg
Benzodiazepines			
Alprazolam, XR	Xanax	0.25–0.5 mg	0.5–2.0 mg
Clonazepam, wafers	Klonopin	0.25–0.5 mg	0.5–2.0 mg
*Lorazepam	Ativan	0.25–0.5 mg	0.5–2.0 mg
*Diazepam	Valium	2 mg	5–20 mg
Hypnotics			
Temazepam	Restoril	15 mg	15–45 mg
Zolpidem	Ambien	5 mg	5–20 mg
Zaleplon	Sonata	5 mg	5–20 mg
Eszopiclone	Lunesta	2 mg	2–3 mg

* Also IV, IM

therapeutic doses. A shorter half-life enhances control during the upward titration process and decreases the risk of accumulation and intoxications.

Mood disorders

Depression in cancer patients requires early recognition and therapeutic intervention. Depression is more challenging to diagnose in patients with cancer because illness produces many neurovegetative symptoms: sleep disturbances, appetite reduction and weight loss, psychomotor retardation, fatigue, apathy, and poor concentration.[16] Focusing the assessment on the psychological symptoms of dysphoria, anhedonia, hopelessness, worthlessness, excessive guilt, and suicidal ideation helps distinguish depression in the context of medical illness.[17]

Table 5.3.7.3 outlines the medically related risk factors for developing depression: increasing levels of debilitation, advanced disease, and concurrent presence of other chronic illnesses or disabilities. Medications frequently encountered in the oncology setting that contribute to depressive symptoms are corticosteroids (dexamethasone and prednisone), chemotherapeutics (interferon, interleukin-2, vincristine, procarbazine, l-asparaginase), and supportive care medications.[17] Depression may relate to organ failure or nutritional, endocrine, and neurologic complications of cancer. Depression is a common symptom of pancreatic cancer, which led to speculation about a tumour-induced mood disturbance mediated by alteration of brain serotonergic function through the effect of proinflammatory cytokines.[18–21]

Management. Psychotropic medications are effective in reducing depressive symptoms present in cancer patients.[22] Table 5.3.7.4 lists the most frequently used antidepressant medications in patients with cancer and their initial and maintenance doses. The antidepressants commonly used today are SSRIs, mirtazapine, venlafaxine, or buproprion. Tricyclic antidepressants and duloxetine are beneficial for patients with depressive symptoms and neuropathic pain. Psychostimulants treat depressive symptoms, and counter fatigue related to advanced illness and the somnolence

Table 5.3.7.3 Medical-related risk factors for depression in patients with cancer

Poorly controlled pain
Other chronic disease/disability; advanced stage
Medications
Corticosteroids
Prednisone, dexamethasone
Inteferon and Interleukin-2
Chemotherapeutic agents
Vincristine, vinblastine, procarbazine, L-asparaginase
Other medications
Cimetidine
Indomethacin
Levodopa
Methyldopa
Pentazocine
Phenmetrazine
Phenobarbital
Propranolol
Rauwolfia alkaloids
Tamoxifen
Antibiotics
(Amphotericin B)
Other medical conditions
Metabolic (anemia; hypercalcemia)
Nutritional (B_{12} or folate)
Endocrine (hyper-hypothyroidism; adrenal insufficiency)
Neurologic (paraneoplastic syndrome)
Sites of cancer
Pancreatic, small cell lung, breast cancer, lymphoma (producing remote CNS effects)

associated with opioids. For depressed cancer patients not expected to survive weeks to months, psychostimulants provide more rapid relief from distressing depressive symptoms. Initiating antidepressants at low doses for elderly and debilitated patients and titration upward as tolerated provides similar benefits, but over a longer period of time.

Suicide and cancer

The incidence of suicide is increased in patients with cancer compared with the general population, but it is not as high as is often assumed. Suicide is more likely to occur in advanced disease as depression, hopelessness, and the presence of poorly controlled symptoms (especially pain) escalate.

Evaluation of suicidal thoughts should take into account disease stage and prognosis. Almost all patients who receive a diagnosis of cancer, even if the prognosis is optimistic, consider or contemplate suicide in the event of developing unbearable or intolerable distress. Some patients maintain supplies of medications for this purpose. This practice allows the patient to maintain a perception of control over progressive disease and feared intolerable pain and inevitable distress. Maintaining this option sometimes allows patients to tolerate difficult treatments.

Morbid preoccupation with suicide or ruminative plans to commit suicide in cancer patients for whom the disease is in remission or in whom a good prognosis exists require careful evaluation.[23] Patients with a poor prognosis, advanced disease, and poorly controlled symptoms often have thoughts of suicide that are more

Table 5.3.7.4 Commonly used antidepressants in cancer

Drug	Brand name	Starting daily dosage PO (mg)	Therapeutic daily dosage PO (mg)
Selective serotonin-reuptake inhibitors			
Sertraline	(Zoloft)	25–50 mg AM	50–150 mg
Fluoxetine	(Prozac)	10–20 mg AM	20–60 mg
Paroxetine	(Paxil)	10–20 mg AM	20–60 mg
Citalopram	(Celexa)	10–20 mg AM	20–60 mg
Escitalopram	(Lexapro)	5–10 mg	10–20 mg
Tricyclics (neuropathic pain management primarily)			
Nortriptyline	(Pamelor)	25–50 mg	50–200 mg
Amitriptyline	(Elavil)	25–50 mg	50–200 mg
Desipramine	(Norpramin)	25–50 mg	50–200 mg
Other agents			
Venlafaxine	(Effexor)	18.75–37.5 mg	75–225 mg
Trazodone	(Desyrel)	50–100 mg	100–200 mg
Bupropion (XL, SR)	(Wellbutrin, Zyban)	50–75 mg	150–400 mg
Mirtazapine	(Remeron)	15 mg HS	15–45 mg
Psychostimulants			
Methylphenidate	(Ritalin)	5–10 mg (8AM & Noon)	10–30 mg
Modafinil	(Provigil)	50–100 mg (8AM & Noon)	100–400 mg
Dextroamphetamine	(Dexedrine)	5–10 mg (8AM &Noon)	10–20 mg

Lithium and mood stabilizers only for bipolar disorder;

MAOIs not recommended.

likely to be viewed as rational by physicians.[24] These patients may request assistance from a physician in obtaining a prescription for medications to use to commit suicide. A treatable major depressive episode may precipitate their suicidal ideation, so it is particularly important to evaluate for the presence of hopelessness, which is a better predictor of suicidal risk than depression itself.[23]

Management. Attentiveness to uncontrolled physical symptoms, especially pain, is crucial. Adequate pain control may have a dual effect of hastening death, while ameliorating suffering. Most physicians feel comfortable providing comfort and relieving distress. Increasing numbers of physicians do not consider this practice to be assisted suicide but as best medical care geared to maximal comfort.[25]

Poorly controlled pain in patients with organ failure and metabolic encephalopathy may result in poor judgement and impulse control leading to unpredictable suicide attempts.[26] These patients benefit from a 24 h companion, nurse, or family member who understands the patient's compromised state and treatment for delirium.

Delirium

Delirium is a global cerebral dysfunction characterized by a fluctuating level of arousal and cognitive disturbances. Symptoms include disorientation and confusion, inattention and poor concentration, perceptual disturbances, disordered thought process, psychomotor agitation or retardation, and an altered sleep–wake cycle. Delirium is distinguished from dementia in part by its reversibility. However, in advanced cancer, as organ failure progresses and results in refractory metabolic derangements, delirium may be irreversible. The primary goal is ensuring the safety of the patient and caregivers. Protecting others from aggressive or combative behaviour is essential. Family members should be told that the cause of the behaviour is brain dysfunction, not a mental aberration, and a given guidance in understanding the patient's states.

In patients with cancer, especially those in advanced stages, an abrupt shift in mood or behaviour is most often related to a change in neurologic, vascular, or metabolic status; a psychological basis is far less likely. In fact, up to three-fourths of terminally ill patients may develop a delirium before death. Common causes of delirium in cancer are outlined in Table 5.3.7.5.

Management begins with attention to the patient's safety. It is important to have constant 1:1 observation, preferably by a person who can correct the patient's misinterpretations of reality. Providing

Table 5.3.7.5 Common causes of delirium in cancer

Causes	Examples
Metabolic encephalopathy because of vital organ failure	Liver, kidney, lung (hypoxia), thyroid, adrenal
Electrolyte imbalance	Sodium, potassium, calcium, glucose
Treatment side effects	Narcotic/analgesics Anticholinergics Phenothiazines Antihistamines Chemotherapuetic agents Steroids Radiation therapy to brain
Infection	Septicemia
Hematologic abnormalities	Microcytic and macrocytic anemias, coagulopathies
Nutritional	General malnutrition, thiamine, folic acid, vitamin B_{12}
Paraneoplastic syndromes	Remote effects of tumors
Metastatic or primary brain tumor	Glioblastoma multiforme, primary CNS lymphoma

Table 5.3.7.6 Behavioral symptoms of delirium in patients with cancer

State	Symptom
Early, mild	Alteration of sleep-wake cycle, transient periods of disorientation Unexplained anxiety and sense of dread Increased irritability, anger, temper outbursts Withdrawal, refusal to talk to staff or relatives New onset of forgetfulness
Late, severe with behavioral changes	Refusal to cooperate with reasonable requests; pulling out tubes and lines Angry, swearing, shouting, abusive Demanding to go home, pacing corridor Illusions (misidentifies staff, visual and sensory clues) Delusions (misinterprets events, usually paranoid, fears of being harmed) Hallucinations (visual and auditory)

consistent caregivers, structured interactions with others, and frequent reorientation may limit the distress experienced by the patient. Elderly patients are more vulnerable to developing delirium, and those with cognitive impairment or dementia are at even higher risk. It is preferable to avoid physical restraints. However, containment of severe agitation may temporarily require restraints to prevent removal of endotracheal tubes, intravenous access, and loss of indwelling catheters, and also to avert falls. Behavioural symptoms of delirium in patients with cancer are outlined in Table 5.3.7.6.

Medications commonly used in managing delirium are summarized in Table 5.3.7.7. Identification and corrections of the underlying aetiology of the delirium is not always possible and in these circumstances, neuroleptics providing relief for the patient from distressing symptoms should be the primary intervention. While haloperidol and lorazepam administered in conjunction provide additional sedative effects for patients with sever agitation,[27] patients with hypoactive subtype of delirium will benefit from the administration of neuroleptics for symptomatic relief.[26,28] In the terminal stage of cancer, delirium may be irreversible and refractory to neuroleptics. Additional sedation with alternative agents may be required to provide comfort and safety for the patient and family.

Table 5.3.7.7 Medications for managing delirium in cancer patients

Drug	Brand name	Approximate daily dosage
Neuroleptics		
Haloperidol	Haldol	0.5–5 mg every 2–12 h, PO, IV, SC, IM
Chlorpromazine	Thorazine	12.5–50 mg every 4–12 h, PO, IV, IM
Risperidone	Risperdal	1–3 mg every 12 h, PO
Olanzapine	Zyprexa	2.5–5 mg every 6–8 h, PO
Quetiapine	Seroquel	12.5–50 mg every 12 h, PO
Benzodiazepines		
Lorazepam	Ativan	0.5–2.0 mg every 1–4 h, PO, IV, IM
Midazolam*		30–100 mg every 24 h, IV, SC
Anesthetics		
Propofol*		10–50 mg every h, IV

PO, orally; IV, intravenously; SC, subcutaneously; IM, intramuscularly.

* Usually IV continuous infusion in intensive care setting.

Further information

www.apos-society.org

Free online education program for multidisciplinary training in Psycho-Oncology. Website contains fifteen web-cast lectures in the five following tracks: Introduction to oncology, program administration, symptom detection and management (eight web-casts), interventions (four web-casts), and population-specific issues.

Direct link: http://www.apos-society.org/professionals/meetings-ed/webcasts/webcasts-multidisciplinary.aspx

www.ipos-society.org

Free online lectures: Multilingual core curriculum in psycho-oncology. Five web-cast lectures translated into English, French, German, Hungarian, Italian, and Spanish.

Direct link: http://www.ipos-society.org/professionals/meetings-ed/core-curriculum/core-curriculum-pres.htm

Holland, J.C. (2002). History of psycho-oncology: overcoming attitudinal and conceptual barriers. *Psychosomatic Medicine*, **64**, 206–21.

Holland, J.C., (ed.) (2006). *Quick reference for oncology clinicians: the psychiatric and psychological dimensions of symptom management*. IPOS Press, Charlottesville.

Psycho-oncology: Journal of the American Psychosocial Oncology Society, Behavioral and Ethical Aspects of Cancer. Published monthly in hard copy and online by John Wiley & Sons, Ltd.

References

1. Cleeland, C.S., Bennett, G.J., and Dantzen, R. (2003). Are the symptoms of cancer and cancer treatment due to a shared biologic mechanism? *Cancer*, **97**, 2919–25.
2. Musselman, D.L., Miller, A.H., Parter, M.R., *et al.* (2001). Higher than normal interleukin-6 concentrations in cancer patients with depression: preliminary findings. *The American Journal of Psychiatry*, **158**, 1252–7.
3. Strain, J. (1998). Adjustment disorders. In *Psycho-oncology* (ed. J.C. Holland), pp. 509–17. Oxford University Press, New York.
4. Hewitt, M., Herdman, R., and Holland, J. (eds.) (2004). *Meeting psychosocial needs of women with breast cancer*. Institute of Medicine, National Academies Press, Washington, DC.
5. Barg, F.K., Cooley, M., Pasacreta, J., *et al.* (1994). Development of a self-administered psychosocial cancer screening tool. *Cancer Practice*, **2**, 288–96.
6. Farber, D.M., Wienerman, B.H., and Kuypers, J.A. (1984). Psychosocial distress in oncology outpatients. *Journal of Psychosocial Oncology*, **2**, 109.
7. Zabora, J. (2001). The prevalence of psychological distress by cancer site. *Psycho-oncology*, **10**, 19–28.
8. Bukberg, J., Penman, D., and Holland, J.C. (1984). Depression in hospitalized cancer patients *Psychosomatic Medicine*, **46**, 199.
9. Jacobsen, P.B., Bovbjerg, D.H., Schwartz, M., *et al.* (1995). Conditioned emotional distress in women receiving chemotherapy for breast cancer. *Journal of Consulting and Clinical Psychology*, **63**, 108–14.
10. Andrykowski, M.A. and Redd, W.H. (1987). Longitudinal analysis of the development of anticipatory nausea. *Journal of Consulting and Clinical Psychology*, **55**, 36.
11. Redd, W.H., Jacobsen, P.B., Die-Trill, M., *et al.* (1987). Cognitive/attentional distraction in the control of conditioned nausea in pediatric oncology patients receiving chemotherapy. *Journal of Consulting and Clinical Psychology*, **55**, 391.
12. Payne, D. and Massie, M. (2000). Anxiety in palliative care. In *Handbook of psychiatry in palliative medicine* (eds. H. Chochinov and W. Breitbart). pp. 63–74. Oxford University Press, New York.
13. Noyes, R. Jr., Holt, C., and Massie, M. (1998). Anxiety disorders. In *Psycho-oncology* (ed. J.C. Holland), pp. 548–63. Oxford University, New York.

14. Rowland, J.H. (1989). Intrapersonal resources: developmental stage of adaptation: adult model. In *Handbook of psycho-oncology: psychological care of the patient with cancer* (eds. J.C. Holland and J.H. Rowland), pp. 25–43. Oxford University Press, New York.
15. Peretz, T., Baider, L., Ever-Hadani, P., *et al.* (1994). Psychological distress in female cancer patients with holocaust experience. *General Hospital Psychiatry*, **16**, 413–18.
16. Roth, A.J. and Holland, J.C. (1994). Treatment of depression in cancer patients. *Primary Care in Cancer*, **14**, 23–9.
17. Coups, E., Winell, J., and Holland, J. (2005). Depression in the context of cancer. In *Biology of depression: from novel insights to therapeutic strategies*, Vol. 1 (eds. J. Lucinio and W. Ma-LeWong), pp. 365–85. Wiley, Weinheim, Germany.
18. Ebrahimi, B., Tucker, S.L., Li, D., *et al.* (2004). Cytokines in pancreatic carcinoma. *Cancer*, **101**, 2727–36.
19. McDaniel, J.S., Musselman, D.L., and Porter, M.R. (1995). Depression in cancer: diagnosis, biology and treatment. *The Journal of General Psychology*, **52**, 89–99.
20. Lerner, D.M., Stoudemire, A., and Rosenstein, D.L. (2001). Cytokine-induced neuropsychiatric toxicity. In *Cytokine therapeutics in infectious diseases* (ed. S.M. Holland), pp. 323–32. Lippincott Williams & Wilkins, Philadelphia, PA.
21. Mussleman, D.L., Lawson, D.H., Gumnick, J.F., *et al.* (2001). Paroxetine for the prevention of depression induced by high-dose interferon alpha. *The New England Journal of Medicine*, **344**, 961–6.
22. Wilson, K., Chochinov, H., de Faye, B., *et al.* (2000). Diagnosis and management of depression in palliative care. In *Handbook of psychiatry in palliative medicine* (eds. H. Chochinov and W. Breitbart), pp. 25–51. Oxford University Press, New York.
23. Rosenfeld, B., Krevo, S., Breitbart, W., *et al.* (2000). Suicide, assisted suicide, and euthanasia in the terminally ill. In *Handbook of psychiatry in palliative medicine* (eds. H. Chochinov and W. Breitbart), pp. 51–63. Oxford University Press, New York.
24. Conwell, Y. and Caine, E.D. (1991). Rational suicide and the right to die: reality and myth. *The New England Journal of Medicine*, **325**, 1100.
25. Chochinov, H.M. (2002). Dignity-conserving care- a new model for palliative care: helping the patient feel valued. *The Journal of the American Medical Association*, **287**, 2253–60.
26. Breitbart, W. and Cohen, K. (1998). Delirium. In *Psycho-oncology* (ed. J.C. Holland), pp. 564–75. Oxford University Press, New York.
27. Breitbart, W., Marotta, R., Platt, M., *et al.* (1996). A double-blind trial of haloperidol, chlorpromazine and lorazepam in treatment of delirium in hospitalized AIDS patients. *The American Journal of Psychiatry*, **153**, 231–7.
28. Breitbart, W. (2002). Spirituality and meaning- centered group psychotherapy interventions in advanced cancer. *Support Care in Cancer*, **10**, 272–80.

5.3.8 Psychiatric aspects of accidents, burns, and other physical trauma

Ulrik Fredrik Malt

Epidemiology of accidents and injury

The one-year prevalence of accidents is about 15–20 per cent with highest prevalence in the younger age groups. About 80 per cent of accidents cause personal injury, and 1/3 to 1/2 of these injuries result in medical attention. About 10 per cent of medically attended injured victims require hospitalization.[1] In the UK (population about 60 million) 31 845 people were killed or seriously injured in 2006 due to road accidents and there were 2 58 404 road casualties.

Accident occurrence and psychiatric disorders

On a group basis, lower social classes, subjects with less education and lower intelligence tend to sustain more accidents and injuries (and have higher morbidity and mortality in general). The ratio of males to females for both fatal and non-fatal accidents is about 2:1 in subjects below 60 years of age. Individual variables associated with increased liability of being involved in an accident include antisocial tendencies, aggressiveness, impulsiveness, thrill and adventure-seeking behaviour. Conscious or unconscious intention is not an important explanation of the overall prevalence of accidents or injuries in the society.

Patients with significant psychological problems (psychopathology including substance abuse) sustain more severe injuries than healthy subjects and the prevalence of psychiatric disorders is increased among hospitalized injured adults compared to surgical patients admitted for other reasons. At least 15–20 per cent of persons brought to hospital emergency rooms due to accidental injury have clinical significant blood concentrations of alcohol. Furthermore, patients with schizophrenia, affective illness and post-traumatic stress disorder have more accidental *deaths* (and suicides) compared to the general population.

Physical injuries

Most non-fatal injuries treated in hospitals are minor head concussion and lacerations, strains/sprains, contusions/abrasions and fractures to body parts such as limbs. More severe injuries are mostly related to high energy accidents (e.g. motor vehicle accidents) and often involve both the head and limbs. Injuries to the inner organs are less frequent, but mostly more severe. The anatomical based Abbreviated Injury Scale (AIS) and Injury Severity Score (ISS) are the most widely used classification system of physical injury. Other classification systems based on physiological impact of trauma (e.g. Revised Trauma Score, Glasgow Coma Score) and combinations of anatomical injury and physiological impact (e.g. Trauma and Injury Severity Score) exist as well. See: http://www.trauma.org/archive/scores/ais.html;

Physical injury as psychological trauma

Accidental injury implies several important sources of threat, loss or conflict which may cause psychological distress or psychiatric disorders. The most important accident related variables associated with subsequent psychological problems include,

- Severity of the accident (e.g. real degree of threat to life of one self and others)
- Degree of helplessness
- Duration of the stressor
- Presence and type of actual physical injury
- Exposure to dead and mutilated bodies.

Nevertheless, pre-accident adjustment, personality and the personal meaning of the accident or injury are the strongest predictors

of both acute psychological responses and long-term psychiatric outcome. This observation holds even in the presence of a severe injury,[2] although the type of injury *per se* may influence the short- and long-term outcome. The relative contribution of 'objective' accident related compared to 'subjective' appraisal related variables in shaping the acute response varies. A rule of thumb is that the less severe the accident, the more important are variables not directly related to the accident *per se* (i.e. the personal meaning of the accident and its consequences for the individual.[3] Important individual variables include,[4–6]

- Pre-injury mental health and adjustment problems
- Personality traits (e.g. neuroticism, quality of attachment)
- Trauma history

The accident *per se* may represent a blow to the person's feeling of invulnerability (narcissistic loss). In some, the accident situation may provoke conflictual feelings (e.g. self-blame, survivor guilt) or shame (e.g. own actions or fantasies prior to the situation). Injury to the body may threaten self-esteem and body image; or represent a loss of function. In some cases, the injury may even serve as a primary gain in a psychodynamic sense. The immediate responses will also be influenced by psychological issues like fear of losing control, or the effect of that phenomenon if it occurs. Conflicts related to secondary gains may also influence the clinical response observed by others.

Clinical features and assessment of trauma at the accident scene

The ABC rule of assessment (Airway, Blood pressure, Circulation) should always be the first step in any medical assessment of acute injury followed by physical examination of the thorax, abdomen, head and finally the extremities. However, except for head injuries associated with impaired cognitive function and injuries that significantly interfere with ventilation or cardiovascular function (e.g. agitation due to hypoxia or apathy due to cardiovascular hypotension), the injury it self plays a minor role for the immediate *psychological* responses to trauma.

Early and marked psychophysiological arousal symptoms like (in decreasing frequency) heartbeat, tremor, dry mouth, restlessness, shaking/trembling, weakness in legs, and sweating are common responses to an acute accident. However, the majority of accident victims appear reasonably calm[5] although many have some degree of inner turmoil that may impair the ability to receive, retrieve, and handle information. If behavioural disturbances are seen during the first seconds to minutes, they mimic phylogenetic responses known from all mammals exposed to acute and severe stressful events: flight, freeze, or fight.

(a) Flight response (anxiety, panic)

The patient appears frightened, may scream or cry. Clear cut panic (e.g. overt confusion, bewildered or aimless behaviour or running away), is rather infrequent even during disasters (<1 per cent). Although lowering of blood pressure is not part of the clinical features of panic, panic is often included in the concept of 'shock' used by lay people and media.

Physiological response to physical injury may be misunderstood as flight response. Patients with injury to the thorax hyperventilate and may appear anxious and scared. Cyanosis is not a sign of emotional distress in adults, and hyperventilation should always be considered as sign of respiratory problems needing urgent medical attention (e.g. pneumothorax). Patients with head injury may be confused and bewildered, but they seldom display the open anxiety seen in patients who panic.

Panic with severe behavioural disturbances may threaten the safety of the subject and provoke anxiety in bystanders and other victims who may themselves be afraid to lose control. Thus, whenever possible, patients with strong anxiety or panic should be offered immediate psychological support. Establishing physical (e.g. hand around the shoulder) and verbal contact is important to reduce panic and provide a sense of security and control. Verbal contact may also reveal the subject's real or imagined fears and provide the subject with an alternative way to express their inner turmoil and despair and thus pave the way for more optimal coping and subsequent behavioural control. The subject should be removed from the accident scene, but not left alone. These subjects need to move around and should not be forcefully immobilized. A helper may walk with the patient until he calms down. The exception to this rule is rare instances where the subject's behaviour is completely out of control representing an immediate threat to the physical safety of self or others.

Reuniting family members may reduce anxiety and worries.

Hyperventilation is treated as usual (breathing into a bag to increase the CO_2 -level) combined with physical and verbal contact as described above. It is crucial that somatic causes (e.g. pneumothorax, intoxication) have been ruled out.

(b) Freeze response (apathy)

Freeze responses include halted surprise or in more extreme cases emotional numbness (apathy). Apathy causing lack of appropriate lifesaving activities occur rather infrequently among random samples of accidentally injured adults (less than 10 per cent). In less than 1 per cent, significant parasympathetic (vagus) responses with lowering of blood pressure occurs ('emotional shock'). These patients appear pale and silent. The look of their eyes gives an impression of detached distance, if they were looking onto their own personal world somewhere far away from the actual accident scene. Rarely, an atypical freeze response characterized by blank denial of having sustained an injury when one, in fact, exists may be seen. These subjects may continue to behave as if nothing had happened and not take appropriate precautions at the accident scene.

Several physical injuries may mimic freeze-response. Patients with internal bleedings (e.g. liver, spleen) may appear pale and silent as if in emotional shock ('freeze response'). The pulse is weak and fast (tachycardia), however, in contrast to the vagus tonus induced bradycardia of the freeze response.

If there is a risk of further injury associated with remaining at the accident scene, patients with freeze responses must be removed to a safe place. They should not be left alone, but covered with a jacket or a blanket over their shoulders and attended to in a calm and gentle way, encouraging them to express some of their thoughts and emotions. If the freeze response is severe and prolonged, the patient should be brought to an emergency room for renewed and extended medical evaluation and basic psychological care. Cases of complete denial of having sustained an injury despite evidence for the opposite, should clinically be handled as a freeze response.

(c) Dissociative symptoms

Dissociative symptoms occur in about 15 per cent during the 1st second to minutes after an accident and may be associated with

flight or freeze responses. Brief symptoms of derealization are most common, even in relatively minor accidents (e.g. 'unreal', like a 'dream' or 'slow movie'). Symptoms of depersonalization (e.g. 'I watched my body burn from a distance') are less common and usually signal a more severe psychological response. Brief symptoms of dissociation do not predict later psychiatric problems,[5,7,8] but marked and prolonged dissociative symptoms still present weeks after the accident.[7]

(d) Fight response (aggression)

Fight responses include irritability, anger and more rarely, open aggression. This response is most often seen among bystanders or helpers who feel threatened by the exposure of dead and mutilated bodies. They may quarrel with the rescue team, and sometimes even interfere with the work of police or helpers. Open aggression is rare among victims themselves with the exception of intoxicated victims with severe personality disorders and a few who have sustained severe head injuries (e.g. subdural hematoma, frontal brain contusion).

Irritability and aggressive comments should not be taken personally by the helpers, but interpreted a symptom of helplessness. In most cases, this response is psychological, but impaired behavioural control due to drug or alcohol may be contributing factors. The patient should be treated as being extremely anxious and under high emotional distress. Reuniting with family or significant others if possible may be helpful. Physical activity may reduce aggression. If suitable, simple tasks which require physical movements may be therapeutic ('Can you give me a hand with . . . . .'), but subjects under stress should never be involved in important rescue tasks due to their impaired judgement ability and tendency to act irrationally.

(e) Acute stress reaction

Marked or severe flight, freeze or fight reponses are included in the ICD-10 (F43.0) definition of acute stress reaction (ASR). ASR is defined as immediate onset of marked psychological symptoms (within 1 hour) following exposure to an exceptional mental or physical stressor. The symptoms must begin to decrease after 8 hours if the stressor is transient (e.g. accident). If exposure to the stress continues (e.g. combat zone, hostage situation) the symptoms must begin to diminish after 48 hours. In contrast, the DSM-IV concept 'acute stress disorder' (ASD) describes development of symptoms not earlier than 2 days after the trauma but within one month after exposure.

Psychotropic drugs are seldom needed to treat acute psychological responses at the accident scene if proper medical care including emotional contact from skilled, empathic helpers is offered. Violence towards victims having lost behavioural control may increase the anxiety among other victims and bystanders, and in fact, increase the risk for more behavioural disturbance within the group, and should thus be avoided.

(f) Acute pain

Some injured persons do not report pain complaints during the 1st second to minutes after even severe physical injury, and some may even continue to perform tasks as usual. This response occurs particularly in situations with continous threat to others or own life (e.g. wounded soldiers). This is part of a brief dissociative response which may be life saving and does not reflect psychopathology. However, a few accident victims respond differently. They may report the most painful physical sensations ever experienced. In the absence of severe physical injury, this response most often reflects catastrophic cognitions associated with severe anxiety[5] and should be treated accordingly. Most injured patients report some degree of pain as minutes pass, however.

Severe pain should be treated at the accident scene and will contribute to psychological and physical recovery from the injury.[9] Anxiety and fear may lead to increased pain complaints, so may imagined (!) severe injuries. For those reasons, it is important not only to examine the presence of actual injury, but also explicitly ask the victim if he or she *believes* or *fears* having sustained serious or life-threatening injuries not detected by the medical personnel. If yes, factual information combined with additional proper medical examination if needed, should be provided to reduce the subject's fears and worries. Faced with true life-threatening injuries, the helper should admit facts if asked, but nevertheless provide some hope and cautioned optimism. It is often hard to evaluate true prognosis at the accident scene and advanced trauma surgery may save the life of many severely injured subjects who would have died a few decades ago.

Responses seen in the emergency room

In urban areas, most subjects will be brought to emergency rooms within less than an hour. At that time most victims have started the process of working through the accident, the injury and its implications. This process is reflected in a characteristic cluster of emotions, cognitions and physiological symptoms observed in humans exposed to all types of stressful situations.[10–12]

- Intrusion includes images of the accident popping into the victims mind, and thinking about the accident even when the person do not want to do so. The main load of intrusive symptoms are related to the severity of the accident and the personal meaning. Intrusive symptoms are common both in post-traumatic depression and anxiety.[5]
- Avoidance includes trying not to talk about the accident or avoiding any cognitive or behavioural activities which reminds the person about the accident. Such symptoms and signs are strongly related to accident-independent variables such as personality traits (e.g. coping style) and more often associated with anxiety than depression.[5]
- Hyperarousal includes startle response, strongly increased heart rate, shivering and trembling, irritability, difficulty in concentrating, hypervigilance and disturbed sleep. With the exception of difficulty to sleep, clinically significant hyperarousal is rather infrequent in randomly selected accidentally injured subjects (less than 10 per cent). However, severe hyperarousal signifies a strong physiological and emotional response and is increased among injured compared to non-injured accident victims and is in some studies associated with later post-traumatic distress problems.[13]

The three most common types of behavioural problems seen in the emergency room are,

- Uncontrolled crying or screaming
- Strong anxiety which may include excessive pain complaints
- Aggression and dyssocial behaviour

Crying and anxiety are associated with high levels of intrusion and avoidance, and may be part of ASR. Systematic and carefully

conducted medical examinations accompanied by supporting questions about the patients emotions, thoughts, and fantasies are the most effective way to put the patient at ease. Sedating drugs are seldom needed if the necessary psychological support is provided. Separation from family members or significant others may increase anxiety and despair, and family reunion may be helpful. If symptoms of high arousal persist, prazosin, a central nervous system (CNS) active alpha-1 adrenoreceptor antagonist or a beta-blocker (e.g. 40–60 mg propranolol) or alfa—may be given to attenuate extreme adrenergic tonus.[14]

Aggressive behaviour occurs in about 5 per cent of injured persons brought to hospital, mostly among intoxicated subjects. The presence of head injury must be ruled out. Most cases can be brought under control with the help of significant others and firm, but calm attitude, addressing the fear or helplessness. In a few cases, acute administration of benzodiazepines or a sedating neuroleptic may be necessary. If the patient is intoxicated or suffer from respiration difficulties, neuroleptics may be the safest option. In cases of armed patients, the necessary precautions must be taken.

Psychotic forms of ASR are seldom seen in injured adults and even patients with schizophrenia or other psychotic disorder prior to the accident appear remarkably calm and collected upon arrival in the hospital. If psychosis is present at arrival in the emergency room, influence of psychoactive substances, severe injury (e.g. brain injury, respiratory failure) or a concurrent psychotic disorder must be ruled out.

(a) Whiplash injury

Rear end collision may cause a whiplash like movement of the neck. Biomechanical studies suggest overstretch of cervical facet-joint capsules as a possible source of pain. Neck pain, stiffness or tenderness may occur minutes to hours after the accident. A medical examination including an X-ray of the cervical columna seldom reveals pathological findings (Quebec classification grade I). In more severe cases, distortion and minor bleeding in capsules, ligaments, tendons or muscles (grade II) may lead to additional musculoskeletal signs such as decreased range of motion and point tenderness. In severe injuries, neurological findings (impaired myostatic reflexes, pareses, loss of sensibility, grade III) or even fractures (grade IV) may be present. In patients with whiplash related injury grade I or II, acute psychological distress and associated neck pain is the most important predictor of long-term outcome.

In the emergency room, treatment should aim at providing the subject with adequate information about the good prognosis. Pain after whiplash-injury usually lasts for four-to-six weeks (!), but gradually disappears. In cases of pain without somatic findings, pain killers or antiflogistic medication have uncertain effect and should not be prescribed for more than a week. Sick leave should be avoided or be as short as possible. Mobilization and early return to work is recommended. Overtreatment by physicians or physiotherapists (e.g. application of stiff collar despite no findings of injury to the cervical columna) may lead to permanent illness behaviour and pain-fixation.[15] The optimal physical treatment of whiplash injury is still unsettled,[16] but premorbid pain and psychiatric disorders represent a risk for development of chronic disabling symptoms and should be treated.

(b) Significant others' needs

Relatives or survivors may want to see dead significant others brought to hospital, and touch them. This process helps the relatives to work through the traumatic event and should be encouraged. If the dead body is grotesquely disfigured, the most horrifying parts should be covered prior to exposure. In any case, a physician or a skilled nurse should accompany the relatives during exposure. Small childrens' emotional response to dead bodies mirror the adults' response. Accordingly, reducing the anxiety and fear of the adults is the best way to help children cope with dead ones. Correspondingly, in cases of severe anxiety in accompanying small children, addressing the helplessness and anxiety of the parents is important. If dead bodies are stored in hospital chapels, care must be taken to cover the presence of religious symbols incongruent to the religious status of the dead one and his family (e.g. Christian crosses should be covered in case of a Jew or a Muslim). The reader is referred to chapter 4.16 for more information on culture specific responses to stress and trauma.

In disaster situations, the need for information varies among relatives, depending on whether their loved ones are missing, injured, or dead (survivor status). Those who have lost loved ones often want to talk to rescuers or get information with regard to any hint about the emotional status of the dead one at the time of death. Accordingly, in situations with several hundred relatives come to the hospital, information is provided in separate groups according to the significant other survivor status. The logistics of such procedures should be outlined in the hospital's disaster plan.

Psychiatric treatment during hospital stay

Most studies indicate that risk factors, emotional, and behavioural responses correspond to that of medically ill patients and identifying those who are at increased risk can follow the same guidelines as for medicine in general.[17] Some patients may complain about physical symptoms suggesting undetected injury. Such complaints may in fact be true. If not addressed and attended to, psychological distress presented by means of somatic complaints or symptoms is the rule.

Clinical syndromes requiring psychiatric attention during hospital stay are listed in Table 5.3.8.1. Complete denial of severity of injury or avoidant coping is maladaptive and should be counteracted.[18] Relatives or significant others should be contacted. They may convey unrealistic fears—or hopes (e.g. 'you will be able to walk'—attitudes in patients with permanent paralysis of legs)—which strongly influence the behaviour and emotional well-being of the patient. They may also provide information which may be helpful to understand current behaviour (e.g. previous dysfunction, 'silent' delirium undetected by staff).

(a) Anxiety and acute stress reaction

Worrying and compulsive thoughts about the accident or the injury (intrusion) is seen both in anxiety and depression. Extreme anxiety may infrequently lead to cardiovascular complications (e.g. pulmonary embolia) in subjects with cardiovascular risk factors (e.g. elderly subjects often smokers with hypertension and arterosclerosis).

Sleep problems may be present or related to physical pain and treatment procedures. The aetiology of nightmares following traumatic injury is complex.[19] They mostly emerge a couple of days after the accident and disappear gradually. Persisting nightmares for more than two weeks without any signs of mastery in the dream content, suggest development of post-traumatic stress disorder and should thus be treated.

Table 5.3.8.1 The most frequently seen psychiatric syndromes during hospital stay following accidental injury

Type of syndrome or clinical problem	Clinical symptoms and signs	Comment
Delirium	Confused, strange behaviour; episodic disorientation; irritability; episodic fearful look	May occur without obvious signs of agitation if sedated; Relatives may detect it and be upset. Diagnosis: 'Draw a clock test' helpful
Abstinence from drug or alcohol	increased pulse; sweating; tremor; insomnia; agitation; anxiety; nausea; abdominal pain; dysphoric.	May be interpreted falsely as accident-provoked anxiety
Antisocial personality, histrionic or borderline personality disorder	Aggressive behaviour; poor compliance with treatment; abusive language; high demand for analgesics	Undetected brain dysfunction must be ruled out; Relatives may provide important pre-injury information
Hypomanic or manic responses	Elated mood; emotions do not correspond to the situation; uncritical behaviour	Undetected brain dysfunction; bipolar disorder or hypomanic response as a defence against survivor guilt
Anxiety	Tense, anxious, restless, worrying, increased startle reflexes; insomnia; dissociative symptoms may occur.	Prolonged or delayed stress response or disorder; imagined or real threat from accident or injury; physical complication (e.g. hypoxia, delirium); abstinent or side effects of drug. Obsessive-compulsive traits and high inner tension with fear of losing emotional control. If unexplained, consult relatives for psychological clues.
Depression	Withdrawn, loss of appetite; inability to feel; sad; worrying; passive; lassitude; anxiety symptoms frequent	Grief; survivor guilt; psychological response to disfigurement or loss (real or imagined) of function or self-esteem; reactivating of previous painful memories
Medically unexplained physiological events including delayed healing of wounds	Rare phenomena; typical senior physicians statement 'I've never seen something like it before'	Secondary gain by extended hospital stay (e.g. alternative prison); factitious disorder; extreme stress (psychophysiological activation)
Excessive pain	Complaints of pain; poor sleep and appetite; poor performance; do not reveal emotions.	Undetected physical complication; insufficient pain treatment; anxiety or depression response in past with obsessive-compulsive traits; withdrawal syndrome; drug abuser.
Partial or complete denial of actual injury	As-if-nothing-has-happened behaviour; refuse treatment; request early dismissal from hospital	Undetected brain dysfunction; psychotic disorder. If male and partial denial, consider obsessive-compulsive traits and high inner tension with fear of passivity

Psychological interventions should be based on clear indication and be brief, distress focused and time limited. Symptoms of intrusion including nightmares may be treated by simple psychological techniques. If one specific traumatic event which can be delineated (e.g. visual image of a traumatic moment), psychological video replay techniques (VRT) may be useful. The subject is taught how to relax. Subsequently, the subject reviews the pre-accident and accident situation on an imagined (i.e. mental) video screen. When the anxiety rises to unacceptable levels, the subject is asked to push the (imagined) stop button and press fast replay until a pre-accident situation where the subjects is at ease is reached. When calm, the procedure is repeated, until the subject can view the whole accident without strong anxiety.

If the anxiety level is high or the traumatic event is more complex, more comprehensive interventions are needed, e.g. Eye Movement Desensitization and Reprocessing (EMDR) or Trauma-Focused Cognitive Behavioural Therapy (TF-CBT).[20] If strong anxiety is not brought under control by means of psychotherapy or other behavioural techniques (e.g. applied relaxation), psychoactive drugs can be added. In injured patients with childhood trauma or other traumatic events in the past, a selective serotonin reuptake inhibitor (SSRI) probably should be the first choice. Sleep problems are best dealt with by environmental adjustment whenever possible, or by optimal pain control if appropiate. Sedative drugs are secondary option. Mianserin or mirtazapine combine sedative and antidepressant effects and may be alternative to benzodiazepines.

Acute stress disorder (ASD) may be conceptualized as an acute form of PTSD and may predict chronic PTSD.[21] The main treatment is psychotherapy (i.e. EMDR, TF-CBT). In PTSD, prazosin reduces nightmares and sleep disturbance in placebo-controlled studies, and may thus be an option also during the first days to weeks after trauma. Some studies have reported propranolol given within days following a traumatic event to be useful for mitigating PTSD symptoms or perhaps even preventing the development of PTSD. The mechanism is thought to be explained by reduced consolidation of emotional memory. Small doses of glucocorticoids may reduce traumatic memories in ASD as well,[22] but larger controlled studies are needed to verify this finding. Benzodiazepines may also reduce acute distress, but may not reduce the risk of 6-month psychiatric anxiety problems. In conclusion, all psychopharmacological treatments must be provided together with psychological interventions addressing the key psychological sources of distress and worry.

(b) Depression

Depressed mood during the first days to weeks following an accident are mostly due to guilt, shame, or grief due to real or

imagined losses. The key to understanding the response is the meaning of the accident or the injury for the patient. Guilt, shame or rumination over real or imagined losses is associated with long-term problems[23] and may require specific therapy.[24] Premorbid causes of depression (e.g. bereavement, mood disorder at the time of the accident), must be kept in mind. In patients with immobilizing injuries staying in hospital for an extended period, some degree of depressive symptoms is the rule. If the symptomatology is severe and persistent, antidepressants are indicated.

Persisting depression has been found to be highly predictive of a long-term psychiatric consequence, and moderate to severe depression predicts less likeliness of returning to preaccident functional level. Thus, depression should always be taken seriously. If marked depression presist for more then 2–3 weeks, an antidepressant should be given. Due to lower incidence of side effects, the newer low-toxic antidepressants are preferred. Delirinm and abstinence are treated as usual. Detailed presentations of the psychopharmacology of the injured or medically ill are available.[34–36]

(c) Pain

Both the injury itself (e.g. burn injuries, injuries to the pelvis, penetrating traumas) or medical treatment (e.g. physiotherapy to prevent contractures, ICU) may be associated with psychological distress and pain.[25,26] Pre-accident psychopathology increases the prevalence and severity of pain complaints. Pain or fear of bringing about pain leads to diminished movement, which can engender contractures, muscle atrophy, and bed ulcers. Traumatic amputation may be associated with phantom pain and exacerbation of pain in response to imagined movements has been reported in subjects with spinal cord injury.[27] Poorly treated pain can be demoralizing to patients and provoke psychological regression, giving-up responses and long-term psychiatric problems[28] including increased risk for suicide.[29] Thus active pain control is crucial and may reduce the prevalence of long-term suffering.[30] Comprehensive reviews of the psychological care of burned subjects are available.[31]

Concerns that trauma patients with injury related pain will become addicted if treated properly with analgesics is neither supported by clinical experience nor by empirical data.[32] However, co-morbid psychiatric disoders must be taken into account when treating injury-associated pain. Patients with a history of substance abuse may have greater tolerance to analgesics and will have to be titrated to higher doses.[33]

(d) Confusion and psychoses

In civilian life situations, confusion or psychotic responses appearing for the first time days after being admitted to hospital are almost exclusively due to a central nervous system dysfunction. Risk factors are severe injuries (ISS >15; third degree burn injury) and major head injuries (e.g. contusion). Impaired cognitive functions or drug or alcohol abuse prior to accident and age >50 increase the risk of organic mental dysfunction.

In a few cases, psychotic-like confusional and agitated responses occurring during the hospital stay may be due to abrupt disruption of intake of psychotropic medication taken for long periods prior to the accident. If in doubt about the aetiology of a psychotic response seen during the first days to weeks (including severe manic episodes), the psychiatrist should consider the response to be of organic origin and explore the pathophysiological processes as done in cases of delirium. Asking the patient to draw a clock, may be a simple and effective way of detecting organic dysfunction.[37]

(e) Alcohol and drug abuse

A significant number of accident victims brought to hospital have an alcohol or drug problem, and symptoms of abstinence may be misinterpreted as psychological anxiety.[38] In cases of grossly deviant behaviour, the presence of co-morbid severe personality disorder should be considered. Alternative explanations include delirium or side effects to drugs (e.g. steroids).

(f) Psychological needs of rescue personnel and staff

Debriefing is a psychological treatment intended to reduce the psychological morbidity that may arise after exposure to accident or injury. Debriefing involves promoting emotional processing/ catharsis or ventilation by systematically encouraging recollection/ ventilation/ reworking of the traumatic event. There is no evidence that this method reduces the incidence of post-accident problems in civilian life.[20,39] Accordingly, psychiatric intervention in the emergency room should be limited to those who display acute and severe psychiatric disorders.

Rescue personnel and medical staff may be psychologically affected by sudden exposure to grotesquely mutilated bodies.[40] The same individual vulnerability found in injured subjects apply. Group debriefing has been recommended if the rescue operation was extremely difficult; there were many dead or there was explicit harsh critique of the rescue operation from the media. Debriefing offered and conducted by a respected senior member of the rescue team is probably more appropriate than debriefing offered by psychologists or psychiatrists. However, empirical data regarding efficacy of emotionally focused group debriefing is scarce.

Long-term behavioural and psychiatric consequences of physical trauma

Physical injury may cause permanent physical change including neurological dysfunction,[41] impaired physical function, changes in perceived somatic health including pain,[30] decreased capacity to work (in children: play), decreased social contact and decreased leisure pleasure.[42] In an unselected population of hospitalized accidentally injured adults, about half will report some complaints three years later. Among those with most severe injuries (ISS >15), only 1/3 will have made full recovery after three years, and about half will report at least moderate disabilities.[43]

The prevalence of non-organic mental disorders among hospitalized adults is about 20 per cent after six months and 10 per cent after two years.[29, 44–49] Depressive symptoms and disorder are most frequently seen followed by specific accident-related phobia and PTSD. Subsyndromal PTSD-cases must be added to these numbers. PTSD is associated with several physical health problems including cardiovascular diseases, respiratory diseases, chronic pain conditions, gastrointestinal illnesses, and cancer.[12] The prevalence of alcohol and drug abuse is increased as well.

The prevalence of long-term psychiatric is increased in injuries associated with visual disfigurement, loss of body parts or physical function (e.g. spinal cord injury), neck injuries and injuries to the pelvis and genital areas.[42] Chronic pain following accidental injury is often associated with concommittant mental disorders, in particular mood disorders or PTSD.[50] Man-made accidental injury

(e.g. assault, combat, rape, terrorism) cause more long-term mental problems than other types of accidental injury (e.g. natural disasters). Studies comparing outcomes in *men versus women* have been mixed. Current evidence suggests that women are at higher risk for anxiety and depression, and men are more at risk for substance abuse and antisocial behaviour.

Following extreme psychological and physical trauma (e.g. torture, concentration camp survivors, hostage situations), permanent change in the person's pattern of perceiving, relating to, and thinking about the environment and the self may occur (ICD-10 F62.0: Enduring personality change after catastrophic experience). The changes should not fully be explained by the presence of PTSD. This diagnostic category does not exist in the DSM-IV.

Assessment of long-term psychiatric consequences of traumatic injury

The following key-points need to be explored when evaluating long-term effects of traumatic injury

- Social and cognitive resources (including social support)
- History of mental disorder, social dysfunction or trauma in the past
- Overlooked physical injury (increased risk if high energy accident or severe injury, e.g. undetected frontal brain damage or other neurological injury)
- Deviant behaviour or accident-related psychiatric disorders following the accident (including ASR or ASD)
- Painful treatment procedures
- Accident-independent traumatic life-events during the post-injury period
- Current psychiatric disorders

Patients, relatives and physicians may evaluate long-term problems differently[51] and psychiatric co-morbidity is prevalent. Thus the clinical assessment should be supplemented by a systematic screening for the most common psychiatric disorders (e.g. MINI neuropsychiatric interview) and cognitive, behavioural and quality-of-life issues (e.g. Impact of Event scale, General Health Questionnaire). Questionnaires specifically designed to address physical, emotional and social outcome of accidental injury are available.[42] A proper, complaint-focused medical examination is often necessary as well.[52] The psychiatrist may improve the quality of the medical examination by providing the examining physician with specific diagnostic questions based on information of the patient's trauma history and symptom complaints.

Treatment of long-term problems

Psychiatric disorders occurring in the aftermath of injury are treated according to general treatment guidelines of mental disorders with some modifications. EMDR and TF-CBT are the best validated psychotherapeutic interventions for trauma related PTSD.[20] If the psychological themes are related to conflicts, family issues or secondary events, short-term psychotherapy as outlined by Horowitz and his group[10] may be conducted. Body-focused treatments may be helpful in some subjects with chronic pain problems after trauma.[53] Randomized controlled treatment trials of accidentally injured adults with post-injury psychosomatic and psychiatric problems are few, however. Comprehensive treatment of patients may provide better results than intervention performed by one single professional only.[54]

Antidepressants should be given in cases of mood disorders or PTSD not responding to psychotherapeutic interventions alone. SSRIs and related drugs are first choice. Drugs acting on nor-adrenalin reuptake alone (e.g. atomoxetine, reboxetin) may increase anxiety and should be avoided. Psychopharmacological treatment of somatoform pain disorders should target both serotonin and nor-adrenalin (e.g. amitryptyline, chlomipramine, duloxetine, venlafaxine). In cases of chronic PTSD with high level of intrusive symptoms, prazosin or propranolol may be added. Betablockers may be valuable as a supplement to anxiety provoking exposure therapy. Benzodiazepines may reduce PTSD-related anxiety, but differences in modulation of skin conductance compared to patients with panic disorder support clinical experience that drug treatment should be supplemented with psychological interventions in order to achieve optimal results. Guidelines for psychopharmacological treatment in patients with co-morbid physical disorders exist (e.g.[36]).

Compensation claims and litigation

Most accidentally injured subjects do not exaggerate their loss,[55] and in non-litigant situations malingering is an unlikely explanation in most cases of chronic disturbances after accidents. Neither is economical settlement followed by significant change in clinical situation in most cases. However, in litigation situations, the patient's problem report may sometimes be exaggerated or even invalid. Studies of personal injury plaintiffs indicate that a significant number report pre-injury functioning superior to that of controls, and malingering has been estimated to 20–30 per cent.

Studies consistently show that delayed-onset PTSD in the absence of any prior symptoms is rare, whereas delayed onsets that represent exacerbations or reactivations of prior symptoms may occur.[56] Untrained subjects are able to endorse symptoms on checklists to meet criteria for diagnoses of major depression, PTSD and GAD, and PTSD self-report measures cannot be used for diagnosis.[57] Furthermore, intrusive symptoms are not PTSD-specific and may be significant in depression as well.[5] This fact is often neglected which explains why some expert testimonies misinterpret depression as being PTSD.

The physician should always try to get patient-independent information from reliable sources (e.g. medical records, general practitioners) also related to pre-injury function[58] before concluding about long-term problems due to physical injury. The possibility that clinically significant brain injury or non-injury related illnesses or psychiatric disorders occurring after the injury, have been overlooked during the medical evaluation part of litigation and compensation cases must be kept in mind.

There is no evidence that physical injury provoke *de novo* bipolar disorder or disorganized schizophrenia, even among severely maltreated subjects.[59] However, permanent injury to frontal and temporal lobes of the brain may provoke manic episodes, paranoid psychoses with schizophrenic-like symptomatology and chronic depression. Both in clinical and court settings, such brain dysfunctions may be overlooked due to lack of classical neurological signs.

Expert testimony should be based on the best available evidence and standards of care, which requires that experts stay current in their field of expertise, and revise old opinions as new information is published. Personal experience alone is rarely sufficient. The psychological difficulties and challenges faced by an expert witness is discussed elsewhere.[60]

Further information

Web links

Physical injury scoring systems

This web link provides access to description and on-line calculation of physical injury trauma scores and links to other resources about physical trauma.

http://www.trauma.org/archive/scores/ais.html

Psychiatric treatment

Several databases providing information about treatment in medicine, including psychiatry and psychosomatic medicine are available. These databases include systematic reviews metaanalysis, clinical trials, and more, including both psychological and biological interventions.

The Cochrane Library

http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME

NICE (National Institute for Health and Clinical Excellence):

http://www.nice.org.uk/).

The limitation of such databases is infrequent updates. In areas with limited research, the conclusions reported may be outdated even shortly after they are published. Thus these databases cannot replace continuous updates from databases of original research like:

PubMed http://www.ncbi.nlm.nih.gov/sites/entrez

PsychInfo.http://www.apa.org/psycinfo/

or National center for post-traumatic stress disorder database with information about treatment of PTSD and traumatic stress for Mental Health Care Providers. http://www.ncptsd.va.gov/ncmain/index.jsp

Other useful web links

Practice guidelines for psychiatric consultation in the general medical settings provided by the

Academy of Psychosomatic Medicine: http://www.apm.org/prac-gui/psy 39-s8.shtml

The European association for consultation-liaison psychiatry and psychosomatics publishes power-point presentations about different aspects of psychiatry in the medical ill or injured: http://www.eaclpp.org/

References

1. Zatzick, D.F., Rivara, F.P., Nathens, A.B., *et al.* (2007). A nationwide US study of post-traumatic stress after hospitalization for physical injury. *Psychological Medicine*, **37**(10), 1469–80.
2. Zatzick, D.F., Grossman, D.C., Russo, J., *et al.* (2006). Predicting posttraumatic stress symptoms longitudinally in a representative sample of hospitalized injured adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry* **45**(10), 1188–95.
3. O'Donnell, M.L., Elliott, P., Wolfgang, B.J., *et al.* (2007). Posttraumatic appraisals in the development and persistence of posttraumatic stress symptoms. *Journal of Traumatic Stress*, **20**(2), 173–82.
4. Krupnick, J.L. and Horowitz, M.J. (1981). Stress response syndromes. Recurrent themes. *Archives of General Psychiatry*, **38**(4), 428–35.
5. Schnyder, U. and Malt, U.F. (1998). Acute stress response patterns to accidental injuries. *Journal of Psychosomatic Research*, **45**(5), 419–24.
6. Creamer, M., McFarlane, A.C., Burgess, P., *et al.* (2005). Psychopathology following trauma: the role of subjective experience. *Journal of affective disorders*, **86**(2–3), 175–82.
7. Murray, J., Ehlers, A., Mayou, R.A., *et al.* (2002). Dissociation and post-traumatic stress disorder: two prospective studies of road traffic accident survivors. *The British Journal of Psychiatry*, **180**, 363–8.
8. Wittmann, L., Moergeli, H., Schnyder, U., *et al.* (2006). Low predictive power of peritraumatic dissociation for PTSD symptoms in accident survivors. *Journal of Traumatic Stress*, **19**(5), 639–51.
9. Saxe, G., Stoddard, F., Courtney, D., *et al.* (2001). Relationship between acute morphine and the course of PTSD in children with burns. *Journal of the American Academy of Child and Adolescent Psychiatry*, **40**(8),915–21.
10. Horowitz, M.J. (1997). *Stress response syndromes: PTSD, grief and adjustment disorders* (3rd edn.). Jason Aronson, Northvale, New Jersey.
11. Skari, H., Malt, U.F., Bjornland, K., *et al.* (2006). Prenatal diagnosis of congenital malformations and parental psychological distress–a prospective longitudinal cohort study. *Prenatal Diagnosis*, **26**(11), 1001–9.
12. Sareen, J., Cox, B.J., Stein, M.B., *et al.* (2007). Physical and mental comorbidity, disability, and suicidal behavior associated with posttraumatic stress disorder in a large community sample. *Psychosomatic Medicine* , **69**(3), 242–8.
13. O'Donnell, M.L., Creamer, M., Elliott, P., *et al.* (2007). Tonic and phasic heart rate as predictors of posttraumatic stress disorder. *Psychosomatic Medicine*, **69**(3), 256–61.
14. Schelling, G. (2007). Post-traumatic stress disorder in somatic disease: lessons from critically ill patients. *Prog Brain Res*, **167**, 229–37.
15. Malleson, A. (2002). *Whiplash and other useful illnesses*. McGill-Queens University Press, Montreal.
16. Verhagen, A.P., Scholten-Peeters, G.G., van, W.S., *et al.* (2007). Conservative treatments for whiplash. *Cochrane database of systematic reviews*, (2):CD003338.
17. Huyse, F.J., and Stiefel, F.C. (2006). Integrated care for the complex medically ill. pp. 1-767. Philadelphia, Saunders. *Medical Clinics of North America*.
18. Dougall, A.L., Ursano, R.J., Posluszny, D.M., *et al.* (2001). Predictors of posttraumatic stress among victims of motor vehicle accidents. *Psychosomatic Medicine*, **63**(3), 402–11.
19. Phelps, A.J., Forbes, D., Creamer, M., *et al.* (2007). Understanding posttraumatic nightmares: An empirical and conceptual review. *Clinical psychology review*.
20. Bisson, J.I., Ehlers, A., Matthews, R., *et al.* (2007). Psychological treatments for chronic post-traumatic stress disorder. Systematic review and meta-analysis. *British Journal of Psychiatry*, **190**, 97–104.
21. Meiser-Stedman, R., Yule, W., Smith, P., *et al.* (2005). Acute stress disorder and posttraumatic stress disorder in children and adolescents involved in assaults or motor vehicle accidents. *American Journal of Psychiatry*, **162**(7), 1381–3.
22. de Quervain, D.J.F. (2008). Glucocorticoid-induced reduction of traumatic memories: implications for the treatment of PTSD. In *Progress in Brain Research* (eds. E.R. de Kloet, M.S. Oitzl, E. Vermetten), pp. 239–47. Elsevier.
23. Kleim, B., Ehlers, A., Glucksman, E., *et al.* (2007). Early predictors of chronicpost-traumatic stress disorder in assault survivors. *Psychological Medicine* , **37**(10), 1457–67.
24. Speckens, A.E., Ehlers, A., Hackmann, A., *et al.* (2007). Intrusive memories and rumination in patients with post-traumatic stress disorder: a phenomenological comparison. *Memory,* **15**(3), 249–57.
25. Richter, J.C., Waydhas, C., Pajonk, F.G., *et al.* (2006). Incidence of posttraumatic stress disorder after prolonged surgical intensive care unit treatment. *Psychosomatics*, **47**(3), 223–30.
26. Berben, S.A., Meijs, T.H., van Dongen, R.T., *et al.* (2007). Pain prevalence and pain relief in trauma patients in the Accident & Emergency department. *Injury*.

27. Gustin, S.M., Wrigley, P.J., Gandevia, S.C., *et al.* (2008). Movement imagery increases pain in people with neuropathic pain following complete thoracic spinal cord injury. *Pain.* **137** (2), 237–24.
28. Norman, S.B., Stein, M.B., Dimsdale, J.E., *et al.* (2007). Pain in the aftermath of trauma is a risk factor for post-traumatic stress disorder. *Psychological Medicine* , **10**, 1–10.
29. Edwards, R.R., Magyar-Russell, G., Thombs, B., *et al.* (2007). Acute pain at discharge from hospitalization is a prospective predictor of long-term suicidal ideation after burn injury. *Archives of Physical Medicine and Rehabilitation,* **88**(12 Suppl 2), S36–S42.
30. Castillo, R.C., Mackenzie, E.J., Wegener, S.T., *et al.* (2006). Prevalence of chronic pain seven years following limb threatening lower extremity trauma. *Pain*, **124**(3), 321–9.
31. Van Loey, N.E., and Van Son, M.J. (2003). Psychopathology and psychological problems in patients with burn scars: epidemiology and management. *American Journal of Clinical Dermatology*, **4**(4), 245–72.
32. Abdi, S., and Zhou, Y. (2002). Management of pain after burn injury. *Current Opinions in Anaesthesiology*, **15**(5), 563–7.
33. Alford, D.P., Compton, P., Samet, J.H., *et al.* (2006). Acute pain management for patients receiving maintenance methadone or buprenorphine therapy. *Annals of Internal Medicine*, **144**(2), 127–34.
34. Simon, A., and Gorman, J. (2004). Psychopharmacological possibilities in the acute disaster setting. *The Psychiatric clinics of North America,* **27**(3), 425–58.
35. Fleminger, S., Greenwood, R.J., Oliver, D.L., *et al.* (2006). Pharmacological management for agitation and aggression in people with acquired brain injury. *Cochrane Database of Systematic Reviews*, **4**: CD003299.
36. Malt, U.F., Llody, G.G. (2007). Psychopharmacological treatment in liaison psychiatry. In (eds. G.G. Lloyd, E. Guthrie), pp. 763–94. Cambridge University Press, Cambridge.
37. Royall, D.R., Cordes, J.A., Polk, M., *et al.* (2006). CLOX: an executive clock drawing task. *Journal of Neurology, Neurosurgery, and Psychiatry*, **64**(5), 588–94.
38. Moss, M., and Burnham, E.L. (2006). Alcohol abuse in the critically ill patient. *Lancet,* **368**(9554), 2231–42.
39. Rose, S., Bisson, J., Wessely, S., *et al.* (2003). A systematic review of single-session psychological interventions ('debriefing') following trauma. *Psychotherapy and Psychosomatics,* **72**(4), 176–84.
40. Benedek, D.M., Fullerton, C., Ursano, R.J., *et al.* (2007). First responders: mental health consequences of natural and human-made disasters for public health and public safety workers. *Annual review of public health,* **28**, 55–68.
41. Gurvits, T.V., Gilbertson, M.W., Lasko, N.B., *et al.* (1997). Neurological status of combat veterans and adult survivors of sexual abuse PTSD. *Annals of the New York Academy of Sciences* , **821**, 468–71.
42. Malt, U.F. (1994). Traumatic effects of accidents. In: *Individual and community responses to trauma and disaster*. (eds. R.J. Ursano, B.G. McCaughey, and C.S. Fullerton), pp. 103–35. Cambridge University Press, Cambridge.
43. Schnyder, U., Moergeli, H., Klaghofer, R., *et al.* (2001). Incidence and prediction of posttraumatic stress disorder symptoms in severely injured accident victims. *American Journal of Psychiatry*, **158**(4), 594–9.
44. Bryant, B., Mayou, R., Wiggs, L., *et al.* (2004). Psychological consequences of road traffi c accidents for children and their mothers. *Psychological Medicine*, **34**(2), 335–46.
45. Creamer, M., O'Donnell, M.L., Pattison, P., *et al.* (2004). The relationship between acute stress disorder and posttraumatic stress disorder in severely injured trauma survivors. *Behaviour Research and Therapy*, **42**(3), 315–28.
46. Fann, J.R., Burington, B., Leonetti, A., *et al.* (2004). Psychiatric illness following traumatic brain injury in an adult health maintenance organization population. *Archives of General Psychiatry*, **61**(1), 53–61.
47. Grieger, T.A., Cozza, S.J., Ursano, R.J., *et al.* (2006). Posttraumatic stress disorder and depression in battle-injured soldiers. *American Journal of Psychiatry* , **163**(10), 1777–83.
48. Glynn, S.M., Shetty, V., Elliot-Brown, K., *et al.* (2007). Chronic posttraumatic stress disorder after facial injury: a 1-year prospective cohort study. *Journal of Trauma* , **62**(2), 410–8.
49. Hoge, C.W., Terhakopian, A., Castro, C.A., *et al.* (2007). Association of posttraumatic stress disorder with somatic symptoms, health care visits, and absenteeism among Iraq war veterans. *American Journal of Psychiatry*, **164**(1), 150–3.
50. Geisser, M.E., Roth, R.S., Bachman, J.E., *et al.* (1996). The relationship between symptoms of post-traumatic stress disorder and pain, affective disturbance and disability among patients with accident and non-accident related pain. *Pain,* **66**(2–3), 207–14.
51. Biddle, D., Elliott, P., Creamer, M., *et al.* (2002). Self-reported problems: a comparison between PTSD-diagnosed veterans, their spouses, and clinicians. *Behaviour Research and Therapy*, **40**(7), 853-65.
52. Frenisy, M.C., Benony, H., Chahraoui, K., *et al.* (2006). Brain injured patients versus multiple trauma patients: some neurobehavioral and psychopathological aspects. *Journal of Trauma*, **60**(5), 1018–26.
53. Haugstad, G.K., Haugstad, T.S., Kirste, U.M., *et al.* (2006). Mensendieck somatocognitive therapy as treatment approach to chronic pelvic pain: results of a randomized controlled intervention study. *American Journal of Obstetrics and Gynecology*, **194**(5), 1303–10.
54. Zatzick, D., Roy-Byrne, P., Russo, J., *et al.* (2004). A randomized effectiveness trial of stepped collaborative care for acutely injured trauma survivors. *Archives of General Psychiatry*, **61**(5), 498–506.
55. Bryant, B., Mayou, R., Lloyd-Bostock, S., *et al.* (1997). Compensation claims following road accidents: a six-year follow-up study. *Medicine, Science, and the Law,* **37**(4), 326–36.
56. Andrews, B., Brewin, C.R., Philpott, R., *et al.* (2007). Delayed-onset posttraumatic stress disorder: a systematic review of the evidence. *American Journal of Psychiatry* , **164**(9), 1319–26.
57. Sumpter, R.E., and McMillan, T.M. (2005). Misdiagnosis of post-traumatic stress disorder following severe traumatic brain injury. *British Journal of Psychiatry*, **186**, 423–6.
58. Duckworth, M.P., and Iezzi, T. (2005). Chronic pain and posttraumatic stress symptoms in litigating motor vehicle accident victims. *Clinical Journal of Pain*, **21**(3), 251–61.
59. Eitinger, L., and Strøm, A. (1973). Mortality and morbidity after excessive stress. Humanities press, New York.
60. Gutheil, T.G., and Simon, R.I. (2005).Narcissistic dimensions of expert witness practice. *Journal of the American Academy of Psychiatry and the Law*, **33**(1), 55–8.

5.4

Obstetric and gynaecological conditions associated with psychiatric disorder

Ian Brockington

Introduction

This chapter covers the psychiatry of menstruation, various manifestations of the desire for children (such as surrogate pregnancy and pseudocyesis), pregnancy and mental health, the psychopathology of parturition, infant loss, postpartum psychiatric disorders, the mother–infant relationship and infanticide.

The psychiatry of menstruation

It has long been realized that menstruation and mental illness are linked. As early as 1827 menstrual mood disorder was used as a defence in filicide.[1] In the 1850s, Brière de Boismont[2] and Schlager[3] carried out the first surveys showing that 20–30 per cent of women suffered a mood disorder before or during the menses—usually irritability or depression, occasionally euphoria. There are descriptions of a wide variety of deviant behaviours, including nymphomania, food cravings, binge drinking, pathological lying, shoplifting, and fire-setting, as well as suicide, violence, homicide, and morbid jealousy. There are other nervous diseases associated with menstruation, including epilepsy, migraine, and hypersomnia.

Recently, there has been much research into the biological basis and treatment of 'premenstrual tension' (or its synonyms). A number of daily rating schedules have been published, but self-devised rating scales, tailored to an individual patient's symptoms, can be used, provided that they are carefully completed every day. Scientific studies are bedevilled by difficulties in defining the disorders.[4] It is not known whether this is one syndrome or many. Irritability is striking, but otherwise the symptoms are common to many other disorders.

Although little is known about the aetiology, progress has been made in treatment. There may be a response to serotonin-reuptake-inhibiting antidepressants (e.g. fluoxetine, chlomipramine). In so far as a luteal-phase defect may be a factor, ovulation-promoting drugs such as clomiphene can be tried. The synthetic steroid danazol, and the gonadorelin agonists (which suppress menstruation), are draconian treatments for severe cases. All interventions should be prescribed in the context of a long-term study using daily ratings.

Rarely, menstruation is linked to a psychosis with acute onset, brief duration, and full recovery. Premenstrual, catamenial, paramenstrual, mid-cycle, and 'epochal' variants have been described.[5] Menstrual psychosis is rare, but perhaps not excessively so. There is a clustering of episodes around puberty and after childbirth, although only a small proportion of menstrual cycles are involved. There are sufficient case reports from Japan, India, and Islamic countries to suggest a worldwide disorder. This is not a specific entity, and most typical examples manifest non-menstrual bipolar disorder at another stage of life. Clinically, it resembles puerperal psychosis. The close relationship between these two psychoses is emphasized by women who develop puerperal and menstrual psychosis at different times.[6] A Japanese investigation showed an association with anovulatory cycles.[7] Pregnancy has a beneficial effect, and there are claims of successful treatment with oral contraceptives, progesterone, clomiphene, danazol, and gonadorelins. The basis for intervention is a long-term study, with a good baseline and exact timing of events in relation to the menstrual cycle.

Infertility

Motherhood is among the strongest and most universal of motivations. For many infertile women, childlessness is the most upsetting experience of their lives, and the yearning for children dominates everything. Infertility is stigmatizing, especially in some cultures. Infertile couples often suffer from self-reproach over sexual indiscretions, abortions, or contraception. They envy fertile couples, and contacts with other people's children, family celebrations, and friends' pregnancies are problematic. The security of the marriage may be threatened by the fear that the spouse will desert to a fertile partner; nevertheless, the marriages themselves are often happy.

Infertility differs from other stresses in its duration. The psychological reaction unfolds over years. When treatment begins, there is a cycle of optimism and hope, with a build-up of tension towards the end of the cycle, followed by disappointment and despair. Sexual functioning comes under strain during the investigation, and the discovery of azoospermia is especially stressful. There is some evidence that stress affects conception, though more prospective studies are needed.

Assisted reproduction

Artificial insemination (using the husband's or partner's semen) has been available from the late eighteenth century, and donor insemination since 1884. Its psychological effects on marriage seem minimal; husbands or partners rarely react with jealousy to the

baby, any more than to an adopted child. The proof that the experience is acceptable is that it is often repeated. One of the principles is privacy, ensuring that donor and couple never meet and remain ignorant of each other's identity. It is felt that violating anonymity might compromise the marriage, since donor and mother are too deeply involved in procreation to regard their relationship with detachment; but times may be changing. The interests of the children have to be considered; donor insemination obscures the genetic lineage, and the child cannot benefit from advances in genetics.

In vitro fertilization (IVF) was first performed in 1978, and was achieved with a donated oocyte in 1984; it is now widely used—in Holland, 1/60 babies are born by IVF. The procedure is harrowing, and counselling is mandatory. There is an increase in multiple births, which are more stressful. But the quality of parenting may be superior to that of families with naturally conceived children.

Surrogate motherhood

This has two meanings:

- A woman is inseminated (artificially or naturally) with the husband's or partner's semen, and surrenders the child to the genetic father and adoptive mother. The surrogate provides oocyte and womb, and is a substitute spouse.
- The wife donates a fertilized oocyte to the surrogate gestational mother. This method, involving *in vitro* fertilization and embryo transfer, is the only way a woman without a uterus can have a child that is genetically her own.

A considerable number of women apply to become surrogate mothers, for motives of financial gain, altruism, pleasure in being pregnant, or atonement.[8] A child can now have 3 mothers—genetic, gestational, and rearing.

Surrogate pregnancy has stirred up an ethical debate. Apart from religious objections, there is concern about the physical and psychological consequences for the gestational mother, and there are endless opportunities for custody disputes and other legal complications. It has been found that the gestational mother does not bond strongly to the foetus, and most surrogate and commissioning mothers do not suffer from psychological problems.

Pseudocyesis

When a woman believes herself to be pregnant and develops symptoms and signs of pregnancy, this is called pseudocyesis. In a classic monograph, Bivin and Klinger[9] collected 444 cases from the literature. Many sufferers were parous, including women with as many as 10 children, and as many as six episodes of pseudocyesis.

The differential diagnosis includes delusions of pregnancy, in which there are no somatic changes. This is a common delusion and can also occur in men. There is also pregnancy simulated for social, mercenary, or legal purposes (e.g. to escape the death penalty).

The clinical features include:

- a firm belief in the pregnancy, usually lasting until the onset of a false labour at 9 months, after which the disorder usually resolves
- amenorrhoea
- morning sickness and/or pica
- enlargement of the breasts and nipples, and even a discharge of colostrum
- abdominal enlargement, caused by muscular contraction, tympanites, fat, or pathological lesions, but without effacement of the navel
- an illusion of foetal movements
- enlargement of the uterus to the size of a 6-week pregnancy.

Modern diagnostic tests have greatly reduced the frequency. The diagnosis should be made on ultrasound examination. Where radiology or ultrasound are unavailable, an examination under anaesthetic is recommended—in the presence of a family member to avoid accusations of abortion.

The psychological basis is usually an intense desire for children, especially in older childless women. In some cases, however, a guilty fear of pregnancy has been the background; this has occasionally led to dangerous attempts at abortion by non-pregnant women. Pseudocyesis is a demonstration of the influence of psyche over soma, mediated by hormonal secretion. It occurs in dogs, cattle, and rodents. Persistence of the corpus luteum would explain breast changes, moderate uterine enlargement, and secretory endometrium; but it is not the only basis: hormonal measurements have been made in at least 30 patients, some of whom had chronic anovulatory states, hyperprolactinaemia, or androgen excess.

These women require psychotherapy. Simply revealing the diagnosis is unsatisfactory because the patient may consult another doctor with the same symptoms, or develop a recurrence. The underlying conflicts must be explored, helping the patient to accept that she is not pregnant.

Sterilization

Women can be prevented from bearing children by various operations on the uterus and Fallopian tubes, indications for which are contraceptive, medical, eugenic, or psychiatric. Sterilization is the most effective and widespread contraceptive method. A large number of studies have looked at its effect on mental health, but many had methodological weaknesses. Ekblad, however, published two thorough studies in 1950s—a general study of 225 women, of whom 99 per cent were interviewed 5 to 6 years later, and a unique study of 60 sterilized women with no living children.[10]

There have been two modern prospective studies. Cooper and colleagues in Oxford[11] interviewed 201 women 4 weeks before non-puerperal tubal sterilization for contraceptive reasons; 190 were re-interviewed 6 months later, and 193 at 18 months after sterilization: the number with psychiatric illness fell from 21 before the operation to 9 at 6 months, and rose to 18 at 18 months. Not surprisingly, the presence of psychiatric disorder before the operation was a predictor of its continued presence; only two who were in good psychological health before the operation developed psychiatric illness 6 months later. A WHO collaborative study, involving five countries (India, Colombia, Nigeria, Philippines, and England), compared 926 sterilized women with 924 who used other methods of contraception: those who chose sterilization had more preoperative psychiatric disorder. The results from the Nottingham field centre[12] were published separately, and found that 9/138 sterilized women had psychiatric disorder before the operation; after surgery there were only three new cases at 6 weeks, and four more at 6 months, less than the control group.

A small minority of sterilized women are troubled by frigidity or severe regret. The most concrete evidence is a request for reversal, several studies of which have been published. Regret is more common in the following groups of women:

- Younger women or those with fewer children: Ekblad[10] found that none of his 60 childless women required hospitalization for depression, but 16 were seriously distressed and 29 expressed a longing for children of their own.
- Those in whom sterilization was the condition for a termination—a barbaric and punitive practice that used to be the rule in some countries.
- Those sterilized at a time of crisis—after parturition, or during a psychiatric illness—when it is difficult to make a balanced judgement.
- Those under external pressure.
- Those with learning difficulties: the issue of sterilization, which has from time to time been practised in various countries, is becoming more important. With a policy of community care and a greater tolerance of sexual activity, there is an increased risk of pregnancy in women with severe learning difficulties, with the spectre of inherited disorders and problems in mothering. Yet these women greatly desire children, and do not have the same resources to compensate for their lack.
- Those sterilized for medical reasons such as inherited disorders, for which medical advances have later provided alternative solutions (e.g. amniocentesis).
- Those who seek sterilization in a context of marital disharmony: after the marriage has failed, the wife may remarry and change her mind about further children.
- Those with religious scruples.

Hysterectomy

This is one of the commonest operations, and is performed in about 10 per cent of women. There have been claims that it leads to 'post-hysterectomy depression'. But this idea has been thoroughly and systematically refuted. Several prospective investigations have shown that mental health improves after hysterectomy. Three comparable Oxford studies, conducted between 1975 and 1990, have addressed this problem: all showed that psychiatric morbidity fell below its preoperative level, or remained low.[13] The ranks of women with 'post-hysterectomy' depression are swollen by those seeking a surgical remedy for psychosomatic complaints.

In younger women, infertility can be a source of discontent. It would not be surprising if the loss of the womb affected feminine identity and libido; but this is probably also a myth. Prospective studies from Oxford,[13] St Louis,[14] and Aberdeen showed an increase in the frequency of intercourse, and of enjoyment. Concomitant oöphorectomy does not adversely affect psychiatric well-being.[15]

The psychiatry of pregnancy

Pregnancy adjustment

The psychopathology of pregnancy needs to be understood in terms of the adjustment all women must make when they conceive. Pregnancy is not only a biological event, but also an adaptive process.[16] A pregnant woman must carry the baby safely through to delivery, and adjust to the sacrifices that motherhood demands. She must ensure the acceptance of the child by the family, develop an attachment to the baby within, and prepare for the birth. She must adjust to the alteration in her physical appearance, and develop a somewhat different relationship with the child's father.

Many pregnancies are unplanned and not initially welcomed. Many women react to conception with grief and anger. A random sample of English mothers showed that 44 per cent of pregnancies were unintentional, including 17 per cent that ended by legal abortion. In married women aged 25 to 29 years with one child, 80 to 84 per cent of pregnancies were planned, compared with 26 per cent in the unmarried.[17] The planning of pregnancy and its acceptance are two different things. The fact of planning does not guarantee acceptance; 6 to 12 per cent of those who plan their pregnancies subsequently regret them. Most unplanned pregnancies are immediately accepted; even if the initial response is negative, gradual acceptance usually follows. In a small proportion of cases, rejection continues to the end of the pregnancy.

Pregnancy has a profound effect on the relationship with the child's father. At every stage this relationship is of the highest importance. A pregnant woman needs increased attention and care and is sensitive to perceived rejection. Pregnancy alters other relationships as well—with the wider family and friends. Many women become closer to their families-of-origin and in-laws.

The change in appearance and shape is sometimes distressing. Some take pride and pleasure in these changes, enjoy the extra attention, and feel an enhanced sense of womanliness. Others are concerned about their loss of figure and facial bloom, weight gain, and stretch marks. Dysmorphophobia, with ideas of reference and social avoidance can ensue.

Pregnancy may be accompanied by medical disorders, and in all there is an interaction between physical and psychological factors. Pica is common, especially geophagia (eating earth or clay), which can lead to iron deficiency anaemia, bowel obstruction, and roundworm infection; other forms of pica can lead to lead poisoning or hypokalaemia. Rarely, hyperemesis can cause Wernicke's encephalopathy, and delirium can complicate chorea gravidarum.[18]

Denial of pregnancy

In women who do not realize they are pregnant, one must distinguish between three different phenomena: unnoticed pregnancy, deliberate concealment, and dissociative denial. A German survey of 29 000 births found 62 women who failed to recognize pregnancy until the 20th week (1/475 births); 12 were not diagnosed until they were in labour with a viable infant (1/2455).[19] A Welsh study obtained similar figures.[20]

The late discovery of an unwelcome pregnancy carries a small risk of suicide. The mother is also at risk of all those complications of delivery that, with modern antenatal care, have become rare. For the child there are increased hazards, including prematurity and neonaticide.

Prenatal attachment

The mother 'bonds' or 'affiliates' to the unborn child in a way analogous to the formation of the mother–infant relationship after birth. Prepartum bonding is catalysed by quickening and probably by ultrasound examination. The mother begins to have fantasies

about the baby and talks affectionately to it. She may engage the husband or partner and other children in 'playing' with the baby. At the same time she prepares for the birth and motherhood ('nesting behaviour').

There is a pathology of the affiliative stage. In some mothers there is minimal attachment even at term. The foetus is viewed as an intrusion, whose movements annoy the mother and disturb her sleep. A poor mother–foetus relationship is one of the predictors of impaired mother–infant bonding. When the mother's attitude to the pregnancy is obstinately rejecting, therapists can direct her attention to the relationship with the child within. Stroking the abdomen and identifying foetal body parts, or telling stories about the baby's future life, have been suggested.

Foetal abuse

When a mother deeply resents her pregnancy, she may try to harm the foetus. This occurs, with determined intent, in self-induced abortion. It may also occur as a manifestation of rage against the baby;[21] a pregnant woman may pound on her abdomen, even to the point of causing bruising.

It is not only the mother who may 'batter' the foetus. Domestic violence is common and may increase during pregnancy, when kicks and blows are directed at the abdomen, rather than the face. The main factors are sexual frustration, substance abuse, jealousy, the mother's irritability, and unreadiness for fatherhood.

The foetus is cushioned from external violence by the amniotic fluid, but can still be damaged by severe abdominal or pelvic injuries. Domestic violence can lead to miscarriage, foetal death, and premature birth. Infants can be damaged by penetrating wounds, and there are over 100 instances of gunshot wounds to the gravid uterus—the result of murderous assaults, attempts to induce a late abortion, or suicide attempts.

Mental illness during pregnancy

(a) Anxiety

For many mothers, pregnancy is a time of considerable anxiety. The first trimester may involve an anguished decision whether to continue or terminate the pregnancy. Those who have previously suffered from prolonged infertility, multiple miscarriages or foetal loss are especially prone to prepartum anxiety. In the third trimester anxiety is centred on three main themes: fears of parturition (tocophobia), of foetal abnormality, and of failure to cope with motherhood.

These anxieties will usually be managed by ventilation and support, but anxiolytic medication can be used cautiously. Of the anxiolytic agents, phenothiazines are relatively safe. Benzodiazepines are contraindicated in the last stages of pregnancy because of foetal intoxication ('the floppy infant syndrome'). Propranolol is best avoided, because of reports of intrauterine growth retardation, and neonatal cardiac and respiratory symptoms.

(b) Depression

Although prepartum depression has not aroused the same interest as postpartum depression, it is no less common. Depression is common in all women in the reproductive age group, and pregnancy is not protective. Depression can be recurrent, and there is an association with puerperal mania.

The frequency of suicide is a vexed question. There are problems about the accuracy of the data since not all suicides are reported to the coroner, not all have necropsies, and not all necropsies include an examination of the uterus. In addition, both suicide and pregnancy are often concealed. One must therefore treat with scepticism those enquiries which do not scrutinize the primary records. Nevertheless, there is evidence that the suicide rate has declined throughout this century; in the first quarter, about 13 per cent of women who committed suicide were pregnant—a rather high figure, suggesting that pregnancy was a risk factor at a time when illegitimate pregnancy was stigmatized. This was confirmed by the thorough mid-twentieth century study of Weir.[22] More recent studies show rates below those in the general population.

Severe prepartum depression is sometimes left untreated, because of fears about the effect of drugs on the foetus. These fears have been exaggerated. No antidepressive drug is known to have teratogenic effects. Most have no effect on the foetus, though fluoxetine may reduce uterine blood flow and paroxetine may cause neonatal pulmonary hypertension. There are reports of toxic effects or withdrawal symptoms in neonates, so that medication is more to be avoided during the last trimester. Electroconvulsive therapy is safe, provided that the mother is competently oxygenated during anaesthesia; pregnant women should be screened for rare syndromes of pseudocholinesterase deficiency before receiving this treatment.

(c) Alcoholism

Pregnancy has a beneficial effect on alcohol addiction, but, if heavy abuse continues, there are severe effects on the foetus. The main effect is retardation of intrauterine growth[23]; although ethanol shortens gestation, the low birth weight is not explained by prematurity, rather the infants are small for gestational age. The infant becomes addicted and may suffer neonatal withdrawal symptoms. Ethanol is also teratogenic, causing 'the foetal alcohol syndrome' (or 'spectrum disorder'), first described in France in 1968.[24] The features include facial dysmorphism due to maxillary hypoplasia, and brain damage, resulting in long-term cognitive impairment and behavioural disorders (see also Chapters 9.2.7 and 10.4). In the detection of these severe complications, systematic prenatal screening for alcohol abuse is useful.

(d) Other addictions

Cannabis is commonly abused by pregnant women; it affects foetal growth, and may lead to long-term neurobehavioural and cognitive deficits. **Lysergic acid diethylamide** may have teratogenic or mutagenic effects. **Phencyclidine** addiction leads to withdrawal symptoms.

Narcotic addicts, like alcoholics, have multiple emotional and social problems, and many do not seek antenatal care. The infants may be affected by maternal malnutrition and infections such as venereal disease, hepatitis, endocarditis, and AIDS. Narcotics are not teratogenic, but a high proportion of the infants are of low birth weight, partly explained by prematurity, and partly by intrauterine growth retardation. A withdrawal syndrome develops in most babies. The perinatal mortality rate and frequency of sudden infant death, are increased. There is an increased incidence of microcephaly, and there may be impaired mental development, although other factors in the maternal life style may account for this. Methadone maintenance reduces the effect on birth weight; but it may depress respiration in the newborn, and lead to a more severe and prolonged withdrawal syndrome, with a greater frequency of seizures. Buprenorphine may be a more suitable maintenance therapy, with milder withdrawal effects. If it is decided to withdraw heroin, this should be done in the second trimester,

replacing it by methadone. Naloxone, which can be given by implant, has been used, although there are concerns about foetal abstinence syndromes. After birth, the infants should be kept in hospital for at least 14 days. Respiratory depression can be treated by naloxone, and seizures and withdrawal symptoms by sedatives such as diazepam, or by tincture of opium.

Cocaine may be teratogenic, causing genitourinary and cardiac abnormalities, but the evidence is conflicting. Its main effects are cardiovascular: it causes uterine vasoconstriction, and this can lead to placental abruption. The infants may suffer cerebral infarction. There is intrauterine growth reduction and an increased incidence of microcephaly. Premature labour is common. There is a withdrawal syndrome, but this is less severe than with narcotics. There is some evidence of an increased risk of sudden infant death. Long-term effects on language development and behaviour are controversial, and may be due to confounding factors such as maternal depression, other drugs, and the environment.

All these mothers should receive close psychiatric supervision and social casework. Hair and meconium analysis improves the diagnosis of opiate and cocaine abuse in mothers who present unexpectedly in labour.

(e) Eating disorders

There are psychological and somatic reasons for an antagonism between pregnancy and anorexia nervosa; nonetheless, most anorexic women recover, and menstruate when their weight reaches about 80 per cent of the standard weight. Ovulation can be induced by clomiphene or menopausal gonadotrophin in those who fail to menstruate. There are numerous case reports and several long-term studies showing that many women with a history of anorexia nervosa give birth to children in the normal way. The overall effect on fertility has been quantified by a 12-year Danish study; the average number of children (0.6) was about one-third the usual figure.[25] The desire for children is shown by the frequency of infertility treatment, planned pregnancy, and breast feeding.

A minority become pregnant while in the throes of the disease. Anorexic amenorrhoea may delay the diagnosis. Pregnancy usually has a beneficial effect; but if the mother continues to restrict her diet, the foetus may suffer from malnutrition. Occasionally it has been necessary to rescue the infant by elective Caesarean section. There is a tendency to relapse in the puerperium. When mothers are actively anorexic, there is often conflict at mealtimes; occasionally children may become involved in their mother's asceticism, and suffer stunted growth.

Bulimia nervosa is often improved by pregnancy. The pressure of the enlarging uterus on the stomach makes bingeing more difficult. About half relapse after delivery.[26] Pregnancy is not much affected by bulimia, but low birth weight has been reported. Bulimic mothers sometimes show deviant mothering, ignoring or excluding their children while overeating or vomiting, or restricting food supplies.

(f) Obstetric factitious disorder

Self-induced illness behaviour can extend into the obstetric domain.[27] Women may induce bleeding to simulate threatened miscarriage, placenta praevia, or postpartum haemorrhage. They may stimulate rupture of the membranes to precipitate an early delivery. Others have been caught manipulating instruments, for example an external tachodynamometer. Two patients even attempted to simulate hydatidiform mole, by adding human chorionic gonadotrophin to blood samples.

(g) Psychosis

Numerous asylum surveys have testified to the lower frequency of psychosis during pregnancy than after delivery. This was confirmed by Kendell and colleagues, in their linkage of Edinburgh obstetric and psychiatric case registers[28]: in a study of 54 087 births, they found rates of 2.1 per month before conception and 2.0 per month during pregnancy, much lower than after childbirth (51 in the first month).

Pregnancy probably has no effect on chronic delusional states, but it does have a beneficial effect on menstrual, bipolar, and possibly cycloid (acute polymorphic) psychoses.[29] Nonetheless, acute manic and cycloid episodes occur during pregnancy, and some seem remarkably similar to puerperal psychosis. They would be regarded as sporadic or random, except that they have been observed in women with a history of puerperal psychosis (at least 13 in the literature).[30] There is an association with multiparity, with the postpartum episode occurring first.

Neuroleptic agents appear to be safe during pregnancy. Phenothiazines and butyrophenones are not teratogenic. The main (but infrequent) hazard is sedation and extrapyramidal symptoms in the newborn. Lithium is relatively dangerous; at least 12 cases of the rare Ebstein's anomaly have been reported. As delivery approaches, reduced renal clearance can result in toxicity with normal doses; eight cases of alarming blood levels (up to 5 mmol/l) have been reported, with coma and convulsions in the mother. Even at normal blood levels, babies exposed to lithium have suffered lethargy, hypotonicity, and other effects. Carbamazepine has been associated with rather high rates of congenital abnormality, and sodium valproate is particularly dangerous, with major abnormalities especially spina bifida, and a foetal valproate syndrome.

(h) Obstetric liaison services

In view of the complexity of the psychological response to pregnancy, and the frequency of anxiety, depression, and other psychiatric disorders, there should be good liaison between obstetric and psychiatric services. In addition to the need to diagnose and treat prepartum psychiatric disorders, the high level of supervision in the antenatal clinics offers an opportunity for preventive psychiatry, by screening for vulnerable women, including those with unwanted pregnancies, severe social problems, or a history of psychosis, addictions, or depression.

The psychopathology of parturition

Childbirth can be one of the severest of human ordeals, and in spite of its brevity, is a time of risk for psychopathology.[18] In advanced countries, all these complications are rare, but may still be common where obstetrics is primitive, or pregnancy denied. Acts of desperation, such as auto-Caesarean section or suicide, and rage attacks, endangering the foetus, are fully described in the older literature. Delirium is well documented; in most cases, it lasts a few hours, starting shortly before delivery and disappearing after the birth, with amnesia for the event; but it can continue into the puerperium, or start immediately after the birth. Engelhard[31] gave the best estimate of its frequency: in a 10-year survey, there were five cases of transitory confusional states in 19 910 births. The existence of this phenomenon aggravates the jurisprudential problem of neonaticide, because, in an unattended delivery, it is impossible

to know whether or not the mother was temporarily confused. Unexplained stupor or coma has also been described during and immediately after delivery.

Infant loss

The child may be lost for a variety of reasons:

- termination of pregnancy at the behest of the mother
- miscarriage, ectopic pregnancy, and late termination of a wanted child for medical reasons
- foetal death *in utero*, stillbirth, neonatal death, and sudden infant death ('cot death', SIDS)
- relinquishment to adoption.

(a) Termination of pregnancy

The indications for abortion include the following:

- medical—to preserve the health and life of the mother
- humanitarian—when pregnancy has resulted from rape or incest
- eugenic—where there is a risk of congenital abnormality
- psychiatric
- social—because pregnancy is untimely and disruptive
- on demand—in the belief that women should be free to decide when to have children.

There has been a debate on the validity of the psychiatric indications; this turns on the psychiatric consequences of a refusal to terminate. Suicide threats are common, but are rarely carried out; nevertheless there can be no doubt that unwanted pregnancy is a factor in completed suicide. A history of puerperal psychosis is not an indication, because it is equally likely to follow abortion; but there are other, arguably more serious, puerperal complications such as mother–infant relationship disorders, which are more common and severe after unwanted pregnancy. These can be avoided by adoption, but the psychological effects of relinquishment are not negligible.

The psychological effects of termination have been thoroughly explored. Most who voluntarily abort suffer no adverse effects, either in the short or long-term. There is often relief, even euphoria, and a reduction in anxiety, depression, anger, guilt, and shame. A minority experience regret and self-reproach over the 'murder' of the baby. Some feel like criminals and worry about punishment, a nemesis of sterility or future congenital malformations. A few develop clinical depression. A Finnish study showed that the suicide rate was increased from $11/10^5$ to $35/10^5$.[32]

There is a literature on 'postabortion psychosis', but both parts of the term have multiple meanings; 'abortion' refers to miscarriage, termination, criminal abortion, and even stillbirth after short gestation, and 'psychosis' includes delirium, Wernicke–Korsakow syndrome, melancholia, and psychogenic paranoid disorders. Manic or cycloid episodes, similar to puerperal psychosis, occur after abortion: apart from epidemiological evidence[33] and individual cases, the association of postabortion and postpartum psychosis in the same woman has been reported on at least 14 occasions.[30] Some episodes occurred after miscarriage, but several followed termination, in most cases performed to prevent a recurrence of puerperal psychosis.

To minimize the psychological risk, prudent decision taking is of the essence, and counselling has a valuable role. The most difficult part of the experience is the loneliness and isolation. Many do not inform their parents, and, when they do, face censure and unwelcome pressure. The attitude of the child's father is crucial. Attempts should be made to involve him in all aspects of the experience; unfortunately his reaction is often unhelpful. It is axiomatic that a woman should make her own decision—one of the most difficult she will ever take. It often has to be taken hastily, in an atmosphere of conflict and turmoil. The best outcomes are found when a woman makes her decision in a context of respect and support from partner, parents, friends, or counsellor.

(b) Miscarriage

This is a common event, perhaps 40 per cent of all conceptions, but only 10 per cent occur after pregnancy is recognized by amenorrhoea or other signs. An ectopic pregnancy is gynaecologically more serious, but has the same psychological effects. The emotional consequences of miscarriage are not trivial, and can be compared to perinatal death—less severe, because there has been little time for attachment to the newly conceived, but still the loss of a greatly desired child. The event itself, with foetal tissue passed suddenly and painfully, may be disturbing. Some of the psychological symptoms may resemble post-traumatic stress disorder, with intrusive re-experiencing ('flashbacks') and nightmares. There is a sense of failure, guilt, and anger. The incidence of depression is four times the rate found in the general population. There may be depressive episodes at the time of the expected delivery, anniversary reactions, and an increased risk of postpartum emotional disorder after a later normal delivery.

Helping a mother who has suffered a miscarriage is a variant of grief therapy, in which her intense distress is shared, and sadness, guilt, and anger ventilated.

Late termination for medical reasons, although a deliberate intervention, is psychologically similar to miscarriage and to foetal death *in utero*. Some wish to continue the pregnancy in the full knowledge that the baby will be abnormal. Depression is common, and grief long-lasting. All these women require counselling, before and after the termination.

(c) Foetal death *in utero*, stillbirth, neonatal death, and sudden infant death

Reactions to these events are generally more severe than to miscarriage, and each has its special characteristics. When the baby dies in late pregnancy, the mother carries a corpse within her, and must undergo a futile labour. If it dies during labour, the loss is sudden and shock pronounced, with a strong sense of unreality. When the child dies in the first week, the parents have to endure great anxiety, with dwindling hope; they may be involved in the decision to switch off the respirator, and witness the child dying. The later death of an infant, when the maternal emotional response is fully developed, especially sudden infant death, is at the very top of the catalogue of calamities; there is no warning or preparation, and the death is followed by a forensic investigation.

(d) Grief after infant loss

This is similar to other grieving, but has its own special character. There is shock, followed by emotional numbness and emptiness, then long-lasting and agonizing sadness. Grief hallucinations (of foetal movements, the baby's face, the infant crying or playing in the cot) may be experienced. There is guilt, anger, and recrimination. There are various crises, especially the disposal of toys, baby

clothes, and nursery furniture, as well as meeting friends and relatives; some, floundering in their embarrassment, are evasive and unable to comfort or sympathize ('wall of silence'). Especially after SIDS, there may be shame, stigma, and even ostracism or malicious speculation. Envy of successful mothers is a problem; there may even be a temptation to steal babies. Surviving children may be confused by their parents' grief, upset by family turmoil, and deprived of attention and care; they are also grieving and preoccupied with their own search for the meaning of death.[34]

When helping the parents,[35] the principles are as follows:

- **Honesty and openness in communication**. The admission of errors is delicate, but the parents' guilt should not be reinforced by the obstetric team's refusal to accept responsibility. Recrimination, litigation, or querulant reactions are common. Staff should accept this as normal, and try not to be defensive. After a stillbirth, most mothers prefer to be segregated, and discharged early. One or more interviews with the consultant obstetrician are indicated. It is essential that the mother is visited by a member of the primary care team. A lactating mother may need bromocriptine, or to donate milk to a milk bank. Hypnotics may help mothers troubled by insomnia. The doctor should be alert for secondary depression.
- **All parents want to know why the baby died**. The necropsy can help, but parents should be warned that often no explanation is found. Necropsies in SIDS are specialized; the pathologist can play a vital psychological role, and should be available for discussion.
- **Mementoes** should be kept, including a photograph. The dignity of naming and a burial ceremony is helpful. The value of seeing and holding the dead baby has been challenged.[36]
- **The bereaved mother needs to share her distress**. A sensitive and sympathetic person, with the time and interest to listen, can help her grieve and accept her loss. This support will often come from the husband or partner, family, or friends. If not, professionals, especially chaplains or nurses, should step in. Self-help groups and voluntary agencies are invaluable for some mothers.
- **The next pregnancy**. No doctrinaire advice can be given about the timing of the next pregnancy. Increased anxiety during pregnancy and the puerperium can be expected.
- **The grieving sibling**. The routine and rhythm of family life should be disturbed as little as possible. The parents should not be afraid to show their emotions—it is best to acknowledge their sadness, and how much they will miss the baby. They should try to give a factual account of what happened, avoiding euphemisms. It is important to reassure the children: they are not responsible and will not lose the love of the parents; neither they nor their parents are in imminent danger of death. The child can be helped to grieve by looking at pictures of the dead sibling, attending the funeral, and visiting the graveyard. (see Chapter 9.3.7 for further information about bereavement in childhood.)

(e) Relinquishment

Adoption used to be the main way to satisfy the longing to rear children and to handle accidental pregnancy, but there has been a great social change in Europe and North America. Since 1950, the number of children born to single mothers has climbed steeply, and is still climbing, but, despite this increase, the number of adoptions is falling steadily. This is not due to spectacular improvements in the infertility treatment, reducing the demand, nor to the relaxation in the abortion laws, reducing the supply, but a new tolerance of single motherhood. In partial compensation for the scarcity of relinquished babies, the practice of adopting foreign-born children has arisen.

Although adoption is on the wane, attention has been focused on the psychological effects of relinquishment.[37] For some relinquishing mothers, giving up the child is a painful, loving act of selfless courage. In others it is the enforced loss of a living child, with a charade of informed consent. Relinquishment is among the most stressful of events. Instead of understanding and support, there is often loneliness and ostracism. Time is no healer; the child continues to exist and can be seen again, and there is often a fantasy of reunion or restitution. As time goes on, there is a new component; the adult child may seek its biological mother, and there is the hope that this event, which she cannot influence, may happen. There has been a growth in the number of organizations to help relinquishing parents find their offspring. Many countries are grappling with the problems of legislating for reunions.

To avoid these severe and prolonged psychological effects, a relinquishing mother needs counselling during the pregnancy. The aim is to emerge from the experience with self-respect and dignity. After delivery the mother should be encouraged to see the infant and photographs should be filed. Follow-up counselling should be continued for at least 6 months. The mother may wish to join a society for relinquishing parents. There is also the relationship with the adopting family to consider. Adoptive parents should accept any gift or token of the natural mother's love. Information on the outcome of the child should be available. Some birth-mothers wish to provide up-to-date information, so that the child knows they are now respected citizens. A recent innovation is the practice of 'open adoption', in which both sets of parents meet. There is even 'continuing open adoption', which means that they remain in contact over the course of the child's development.

The psychiatry of the postpartum period

The normal puerperium

For many or most mothers, giving birth is a supreme moment, and euphoria or elation is common. Some may be too excited to sleep. These feelings of peace, fulfilment, and accomplishment help to sustain mothers during the weeks of strain that follow. Prolonged euphoric reactions, lasting a week or more, are probably mild puerperal mania, and are often followed by depression.

Newly delivered mothers have to face a number of challenges, including the following.

- **Physical exhaustion**. This can be coupled with the painful sequelae of pelvic trauma.
- **Breast feeding**. Although this has many advantages, it is often difficult to establish.
- **Insomnia**. Sleep deprivation, especially during the first month, is a cause of irritability, and should be borne in mind when mothers present 'at the end of their tether'.
- **Recovery of normal figure and attractiveness**. This may be threatened by weight gain and stretch marks. Mothers may occasionally develop a state similar to dysmorphophobia.[38]

- **Loss of libido**. Episiotomy and vaginal trauma often cause dyspareunia; fatigue may depress sexual activity. Nevertheless sexual relations are usually resumed within 1 to 3 months, though reduced in frequency, and with a delayed return of orgasm. For this and other reasons (e.g. jealousy) the marriage may come under strain.
- **Social privation**. The loss of employment, income, and leisure, as well as confinement to the house, are all contributory factors.

With this background of rapid biological, social, and emotional transition, it is not surprising that a wide variety of psychiatric disorders occur; indeed the psychiatric complications of childbirth are more numerous and complex than in any other human situation.

The **maternity 'blues'** is so common as to be almost normal. Usually between the third and fifth days, many mothers experience a sudden, fleeting, and unexpected period of sensitivity and uncharacteristic weeping. In the great majority this passes off within a few hours, or a day or two. There is some evidence for an association between this brief dysphoric reaction and postpartum depression.

Reactions to severe labours

(a) Post-traumatic stress disorder

After excessively painful labours, some women suffer nightmares, and repetitive daytime intrusion of images and memories, similar to those that occur after the harrowing experiences of war and natural disaster. Since the original description,[39] over 40 papers have been published on this subject, including 10 quantitative studies showing rates of up to 5.9 per cent of deliveries. Many of these women avoid further pregnancy (secondary tocophobia), and those who become pregnant again may experience a return of symptoms, especially in the last trimester.

This disorder can be treated by counselling and by specific psychological therapies. Tocophobia is an indication for elective Caesarean section.

(b) Querulant reactions

Another reaction to a severe labour experience is pathological complaining (*Querulantenwahn*). These women complain bitterly about perceived mismanagement, and their angry rumination may continue for weeks or months, interfering with infant care. Some confine themselves to vengeful fantasies and verbal or written criticism, but others proceed to litigation. Careful assessment is needed to distinguish these reactions from reasonable complaining.

This disorder can be treated by a psychotherapeutic approach, which distracts the mother from her grievances and reinforces productive child-centred activity.

Postpartum anxiety disorders

Recent research has shown that postpartum anxiety disorders are just as common as postpartum depression.[38,40] A review of eight studies of 'panic disorder' showed that 44 per cent of anxious women had an exacerbation, and 10 per cent a new onset, in the puerperium.[41] It is important to identify the focus, as well as the form, of anxiety, because there are several themes that indicate specific psychological therapies. Benzodiazepines should be used with caution in lactating mothers. They are well absorbed from the gut, and more slowly metabolized in the neonatal liver, and occasionally cause lethargy and weight loss in breast-fed infants.

(a) Puerperal panic, and phobic avoidance of the infant

Some mothers, especially *primiparae* in isolated 'nuclear' families, are overwhelmed by the responsibility of caring for the newborn.[42] The panic and agitation seen in extreme examples is an exaggeration of the anxiety that many women experience when they first confront this awesome task. If no help is available, a mother can develop a phobic avoidance of the infant,[43] and risks losing her mothering role.

These disorders can often be handled by the wider family, without invoking professional help. The mother needs sedation, especially at night. During waking hours, she should remain with the baby, but must be supported at all times. Treatment is by desensitization. Gradually the mother takes over, at her own pace, undertaking the easiest tasks first, and involved in all decisions. In severe cases, conjoint admission may be the only way to rescue the situation. With correct diagnosis and management the prognosis is excellent.

(b) Anxieties about infant health and survival

The care of an infant involves ceaseless vigilance. In women prone to anxiety and excessive worrying, or in those who have suffered years of infertility or recurrent miscarriage, motherhood can lead to excessive solicitude about banal tasks that put the baby at risk (e.g. bathing) and sensitivity to the slightest indication of illness. In some, the anxiety is focused on the possibility of sudden infant death.[44] These mothers lie awake listening to the baby's breathing; sleep with their hand on the infant's chest, check the infant many times each night, or even wake the baby to ensure that he or she is still alive. This results in excruciating tension, insomnia, and exhaustion.

These mothers require anxiety management. Day-hospital attendance, with relaxation therapy and group support is ideal. A mother with 'fear of cot death syndrome' may be helped by explanations about the rarity of SIDS, and the infant's resistance to asphyxia, as well as devices to monitor the infant's breathing. The vicious cycle of insomnia and hypervigilance can be interrupted periodically by involving relatives or friends, so that she can sleep under sedation. Ventilation, and the support of mothers who have recovered from similar problems, is helpful. However, these are only palliatives, because the underlying cause is an event, which, albeit uncommon, remains possible during a period of several months.

Puerperal obsessional disorders

There is evidence that the puerperium is one of the main precipitants of obsessive–compulsive disorders.[45,46] In addition to obsessional rituals, the disorder may present with thoughts, images, or impulses of child harm. These impulses to attack the child must be distinguished from the pathological anger that precedes child abuse. The mother is gentle and devoted. She experiences extravagant infanticidal images, such as stabbing, decapitation, or strangulation. She fears being left alone with her infant, and may take extraordinary precautions.[47] The obsessional content may be of child sexual abuse, for instance masturbating or castrating their sons.

Ventilation, explanation, and psychotropic medication are part of the treatment, but are rarely sufficient. It is important to discourage avoidance of the child, and encourage cuddling and play, thus strengthening positive maternal feelings. Cognitive–behavioural treatment can help her to achieve mastery over

irrational impulses. (For the treatment of obsessive–compulsive disorder, see Chapter 6.3.2.1.)

Depression

Puerperal melancholia was one of the first postpartum psychiatric disorders to be identified. During the asylum era, only the most severe cases were admitted, and the occurrence was underestimated when compared with 'puerperal mania'. When, in the 1950s, attention turned to milder disorders, postpartum depression was found to be common in the general population. The pioneering work of the Gordons in New Jersey[48] was soon widely confirmed. In the last 10 years there has been a flood of papers from all over the world. Surveys have shown rates of at least 10–20 per cent, or even higher in the 'Third-World'.[49] 'Postnatal depression' has become a household word. It is an important lay concept, which has legitimized maternal depression in the minds of the public, providing a valid explanation for role failure, diminishing stigma, enabling mothers to accept that they are ill, and to come forward for treatment. It is a slogan that can be wielded in the political struggle to obtain better services for mothers of young families. There is a need for such concepts, which have social influence.

However, one must examine the scientific value of this concept with scepticism. Depression after childbirth is clinically similar to any other depression, and the association of depression with the puerperium is not striking. Whatever the prevalence in surveys, only about 5 per cent consult their general practitioners. The epidemiological evidence is weak. The suicide rate in the first postpartum year is below the female rate.[32] Depression is common in women during the reproductive years, whether they are infertile, pregnant, puerperal, menopausal, or involved in child rearing. The term 'postnatal depression' has the danger of introducing into the minds of the unwary the mirage of a homogeneous disorder with a single cause. Rather, it is a rubric for a heterogeneous group of disorders. Many mothers with anxiety, obsessional, or post-traumatic disorders, or with a disturbed infant relationship, are depressed, but the setting, causes, and treatment are different. Not surprisingly, research into its causes has found that they are the same as those that cause depression at all ages—heredity, a history of previous or prepartum depression, 'neuroticism', adverse events or social conditions, difficult relationships, and social isolation. It has been suggested that the burden of child rearing, rather than child bearing, is a factor, but this has been challenged by a Swedish twin study[50] and a Norwegian suicide study[51] showing that parous women have a lower risk of depression than nulliparous women. In mothers with recurrent puerperal depression one would expect to find specific factors. Adjustment to motherhood has received much less emphasis than it deserves; unwanted pregnancy has been found to be a predictor of antepartum and postpartum depression.

Whatever its frequency, the effects of depression on family life, and the emotional climate in which children are reared, is of great concern. A growing child needs emotional support, attention, approbation, and stimulation. The mother is the child's primary environment, and her mood dominates his or her world. Even very young infants are disturbed by deviant social behaviour in the mother. Although deficits are not universal,[52] her depression can lead to inattention through anergia or brooding, reduced quantity, quality, and variety of interaction, and loss of the reinforcement of the mother's gaiety and tenderness. Her anger may be misdirected at the children. Frequent irritability, impatience, and criticism induce social withdrawal, anxiety, and reciprocal anger. There may be educational deficits. These effects depend on the degree and duration of maternal depression, and the extent to which it involves interactions with the child. (See Chapter 9.3.6 for further information on the effects of maternal depression on child development.) In extreme cases, maternal depression can lead to the tragedy of combined suicide and filicide.

Treatment begins with effective diagnosis. Many more mothers are depressed than ever make their way to the surgery. The reasons for the failure to seek help are not fully understood: some recover early, some do not realize they are ill, and some are ashamed of confessing their symptoms, suffering in silence because of ignorance, stigma, and fears of losing their baby. Screening procedures help the primary care services to identify cases; an example is the Edinburgh Postnatal Depression Scale,[53] which has high sensitivity and specificity. Patients identified by screening, or self-referral, require a full psychiatric examination, in order to identify vulnerability factors and the specific components of postpartum disorders. This initial interview is best held at home, because clinic attendance is an obstacle for mothers fettered with the care of young children, and because domiciliary assessment has a quality that cannot be achieved in the office. The interview should explore the symptoms and course of the illness, study its context in the mother's life history, personality, and circumstances, review the events of this pregnancy, explore the mother's relationships with her spouse, baby, other children, and family of origin, and establish the available supports.

Treatment is focused on depression and any underlying vulnerability. It will always involve psychotherapy, if only in the form of a single interview; it will usually include medication or other specific treatments; a few require electroconvulsive therapy. Working with the baby's father, potentially the main supporter, is important; fathers can come under strain, either because their wife's intimacy with the baby disturbs conjugal dynamics, or because her depression has a domino effect on him. Home visits by community nurses are an ideal method of delivering continuing care and psychotherapy. An extensive literature has accumulated demonstrating the efficacy of psychological treatments by double-blind randomized controlled trials. As for drug treatment, there is no evidence that any drug is superior to others. There are at least 50 reviews of drug treatment in lactating mothers. The suckling infant has little body fat, less plasma protein-binding, an immature liver and kidney and an undeveloped blood-brain barrier. But the risks are minor. Only a minute dose is delivered to the infant: Epperson and colleagues demonstrated that serotonin re-uptake blocking agents do not affect serotonin levels in breast fed infants.[54] Occasionally, babies have been over sedated. It is not recommended that antidepressant agents be withheld, or that breast-feeding be stopped; but it is wise to use these drugs cautiously, and it may be helpful to take the drug after breast feeding (see also Chapter 6.2.3).

Prevention is important in mothers with a history of severe or prolonged postpartum depression. They often present during the next pregnancy, requesting advice or prophylaxis. If they are already symptomatic, or have obvious risk factors such as marital friction or social isolation, they need support from community psychiatric nurses, voluntary agencies, or other groups. If they are well, it is only necessary to establish contact, so that a recurrence is diagnosed and treated promptly. Prophylactic antidepressant medication can be considered.

Mother–infant relationship disorders

Just as the emerging relationship with the foetus is important during pregnancy, so also the growth of the mother–infant relationship is the key psychological process in the puerperium. 'Bonding' is a popular lay term; some professionals prefer 'attachment', but one must not confuse this with infant–mother attachment. The mother–infant relationship consists essentially of ideas and emotions aroused by the infant, which find their expression in affectionate and protective behaviour. Its immense power is revealed in self-sacrifice, and the pains of separation. Its inner presence is betrayed by external signs—touching and fondling, kissing, cuddling and comforting, prolonged gazing and smiling, baby talk and cooing, recognizing signals, tolerating demands, and resisting separation; but it is hard to select a single activity that lies at the core. Particular behaviours wax and wane, but the relationship endures, even when the child is absent, even when it is gone for ever. This emotional response enables the mother to maintain the never ending vigilance, and endure the exhausting toil of the nurture of the newborn.

There is no 'critical period' in the development of the maternal response. Close proximity from the start ('rooming-in') gives confidence in mothering skills, and breast-feeding may help. The infant plays an important part. At an early stage, it can discriminate speech, and reacts preferentially to the human face and voice. Eye-to-eye contact mediates the interaction, and gazing becomes an absorbing activity on both sides. The baby's smile is another catalyst. Videotape studies have shown the infant contributing to a dialogue with its caregiver. Sometimes the maternal response is immediate, primed by affiliation to the foetus, but sometimes there is a worrying delay. For the first 3 to 4 weeks many mothers feel bruised, tired, and insecure, and their babies seem strange and distanced. As the baby begins to respond socially, a normal relationship develops rapidly.

The term 'mother–infant relationship (or bonding) disorders' covers a spectrum of clinical states, which has two main dimensions:

- An absent or negative emotional response. In severe cases, the mother regrets the pregnancy and expresses dislike or hatred of her baby. She may try to persuade her own mother, or another relative to take over, and may demand that the infant be fostered or adopted. The most poignant manifestation is a secret wish that the baby 'disappear'—be stolen, or die.
- Pathological anger. The infant's demands anger the mother and provoke aggressive impulses, which may lead to shouting, cursing, screaming, or assaults.

There is at present little data on the frequency of these disorders in the general community. At the level of 'threatened rejection'—where the mother has an aversion to her child and seeks temporary escape from child care (the threshold for active intervention)— the frequency of these disorders in the general community is probably about 1 per cent. It is much higher in mothers who seek help for 'postnatal depression'—about 10 per cent at the level of established rejection, and another 15 per cent for threatened rejection.[55]

These disorders are usually accompanied by depression, but there are many reasons to reject the euphemism 'postpartum depression with impaired mother–infant interaction'.

- A disturbed relationship is different from a mood disorder.
- When depression is associated with phobias, obsessions, or deviant behaviour, these co-morbid phenomena are still considered worthy of study and treatment in their own right.
- Impaired interaction, although it can be recorded and measured, is not the essence of the phenomenon, but merely its behavioural manifestation; it has other causes, especially infant-centred anxiety.
- Aversion to the infant is not confined to depressed mothers;[56,57]
- When they coexist, their severity and course often differ.
- Only a minority of depressed mothers have this problem. It is important to select them for treatment and not to stigmatize the others.
- The risks—including child abuse and neglect—are higher. It is these disorders, rather than uncomplicated depression, that have serious and long-term effects on the child's development.[58] These mothers are a high-risk group that can contribute to the vital task of preventing child abuse and neglect.
- The treatment is different and specific (see below). Assessment and treatment of this relationship forms an important part of the work of mother–infant mental health teams.
- Management must be aware of the need for training and service provision.
- The aetiology is different, with more emphasis on unwanted pregnancy and challenging infant behaviour.

The diagnosis can be facilitated by self-rating questionnaires, but the main clinical resource is an interview probing the mother's emotional response and behaviour. In severe cases and in research, the 'gold standard' is direct observation, preferably over a substantial length of time.

The treatment proceeds in stages:

- Where there is a delay in the maternal emotional response, explanation and reassurance are usually sufficient.
- When hostility, rejection, and anger are prominent, the primary decision is whether to attempt treatment or not. The mother must be given freedom of choice; it is dangerous for her to feel trapped in unwelcome motherhood. At the same time, the father has his rights. The option of relinquishing the infant must be openly acknowledged, and fully discussed with both parents.
- If it is decided to embark on treatment (as in most cases), depression should be treated with psychotherapy, drugs, or (occasionally) electroconvulsive therapy.
- The specific element of therapy is working on the dyadic relationship. This relationship, like others, grows through shared pleasure. The baby alone has the power to awaken its mother's feelings, so the aim is to create circumstances in which mother and child can enjoy each other. It is a mistake to separate the mother and baby completely, which merely compounds the problem by adding an element of avoidance. If there is any hint of abuse or aggressive impulses, the mother must never be left alone with her infant. She must be relieved of irksome burdens of infant care. When mother and baby are calm, she is encouraged and helped to interact with him—to cuddle, talk, play, and bring out his smile and laughter. Participant play therapy and baby massage may assist.

Treatment can take place in various settings. Home treatment can be successful, provided there is enough support to relieve the mother of night care and stressful duties: the maternal grandmother,

an understanding husband or partner or a family group can sometimes achieve this. Day-hospital treatment provides individual support and group discussion, as well as specific therapies. In the most severe and refractory cases, the proper setting is an inpatient mother-and-baby unit, where an experienced team of psychiatric and nursery nurses, available 24 h a day and 7 days a week, can provide full support. Even in the most severe cases, one can feel optimistic about a successful outcome. (For further information about child abuse see Chapter 9.3.3.)

Postpartum psychoses

These fall into two main groups—organic psychoses and bipolar disorders; a third group—reactive or psychogenic psychosis—is most convincingly seen in adoptive mothers and fathers. There are many causes of delirium after childbirth.[18] Organic psychoses are hardly ever seen in Europe or North America, but may still be important in Africa,[59] India, South-East Asia, and Latin America, where the majority of children are born. Historically the most important causes have been infection and eclampsia psychosis, but cerebral venous thrombosis is common in India.[60]

The form of psychosis still seen in Europe and North America was described by Osiander,[61] and illuminated by the case studies of Esquirol.[62] The long-standing controversy about its nosology has been resolved in favour of a relationship with the bipolar group; but there is also a connection with acute polymorphic (cycloid) psychosis, which may also belong to the bipolar rubric. These are biological brain disorders, with high heritability and an inborn tendency to develop episodes throughout life. The problem of causation can be broken down into three subsidiary questions: the nature of the diathesis, the determinants of clinical polarity (mania, depression, or cycloid), and the trigger that provokes the episode. The first two questions belong to the wider study of bipolar disorder. The third is specific to puerperal psychosis. The clinical facts suggest not one, but several triggers related to the female reproductive process—abortion, pregnancy itself (especially the last trimester), the early puerperium (especially the first 10 days), postpartum menstruation, menstruation in general (see menstrual psychosis above), and weaning. These triggers can be added to the list of other biological events that trigger bipolar episodes, including surgery, adrenocortical steroid treatment, and seasonal climatic changes. Instances can be given of the combination of all these triggers in the life history of individual women, and there may be a shared pathway in these diverse precipitants. The incidence is somewhat less than 1 in 1000 pregnancies.[28,63] The Edinburgh study showed no link with twin pregnancies, breast feeding, single parenthood, or stillbirth. The miscarriage rate has been low in two studies. This psychosis has high heritability—not just for bipolar disorder, but also the puerperal trigger.[64]

Postpartum bipolar psychoses are acute, rapidly reaching a climax of severity. The onset is usually between 2 and 14 days after delivery. Mania is severe, often with 'schizoaffective' symptoms or extreme excitement. Almost every psychotic symptom may be seen—the whole gamut of delusions, verbal hallucinations, disorders of the will and self, and catatonic features. There is often an apparent confusion or perplexity. Since the advent of electroconvulsive therapy and neuroleptic medication, the duration has fallen to a few weeks. A minority of patients show a tendency to relapse in rhythm with menstruation. Puerperal recurrences occur after 20 to 25 per cent of subsequent pregnancies. Non-puerperal recurrences are also common.

There are no specific treatments. The first resource is sedation by neuroleptic agents, but these should be used with caution because of the risk of severe extrapyramidal side effects, which include neuroleptic malignant syndrome. It is usual to stop breast-feeding, although this may not be necessary, because the infant receives only a minute dose of the neuroleptic and adverse effects have not been noted; Clozapine may, however, accumulate in breast milk. Lithium has been used increasingly since the link with manic-depressive psychosis was recognized. There is also evidence for its prophylactic value in women at high-risk. However, it may have adverse effects on breast-fed infants. Electroconvulsive therapy is highly effective in all varieties of puerperal psychosis, including puerperal mania. The location of treatment is an important issue. Since hospitalization can be disruptive to the family, this disorder should for preference be treated at home, where the patient can maintain her role as wife, homemaker, and mother, and her relationship with the newborn; but its severity and the lack of community resources make this a counsel of perfection. If hospital admission is necessary, there are great advantages in conjoint mother and baby admission.

Services for mothers with mental illness

An outline of the services required for this area of psychiatry is slowly emerging. Its aims are prevention in those who are vulnerable, early and accurate diagnosis, and rapid effective intervention, with minimal disruption of family life. These aims require the following:

- **A multi-disciplinary specialist team** This team, a key resource whatever the cultural background, should consist of psychiatrists, nursing staff of various kinds, psychologists and social workers; it can serve a population of several million inhabitants, handle severe and intractable illness, train all staff, develop services, and conduct research.
- **A community service** Domiciliary assessment and home treatment are appropriate for mothers.
- **Day care** A day hospital can provide a full range of interventions, including groups, play therapy, motherhood classes, anxiety management, and occupational therapy, with minimal family disruption. The presence of mothers with similar disorders is an additional support. The children are cared for in a crèche.
- **Inpatient facilities** Conjoint admission of mother and infant is superior to the admission of the mother alone.[65] Wards dedicated to conjoint admission also have advantages over admissions to general psychiatric wards, although they are more expensive.
- **An obstetric liaison service** Apart from treating prepartum mental illness, this provides an opportunity for preventive psychiatry, by detecting vulnerability during pregnancy.
- **Links with other agencies providing services for mothers** The social services have a key role. Their family centres fulfil a similar function to mother-and-baby day hospitals. They can relieve the burden on the mother, and safeguard the child, by providing emergency foster care. Other agencies include the National Society for the Prevention of Cruelty to Children (in the United Kingdom), midwifery services, primary care teams, and child psychiatry services.

- **A network of voluntary organizations** These are independent organizations, but can have close cordial ties with the professional service. There is no one better suited to support a depressed mother than another mother who has suffered a similar problem and is now well—she knows the stratagems or words of comfort that were helpful, and is living proof of the hope of recovery. For each disorder, a panel of recovered mothers is an important resource.
- **Medico-legal expertise** Expert advice is often required in cases of child abuse or infanticide, and where a mother with mental illness is seeking custody of, or access to, her children.

The psychiatry of parental children-killing

The term 'infanticide' covers the killing of infants and children by their mother or father in a wide variety of circumstances, broadly divided into 'neonaticide' (killing the newborn) and 'filicide' (the later murder of a child).

Neonaticide

Killing neonates (especially female infants) has been customary in certain societies as an official policy or 'grass-roots' custom for controlling population growth. This is completely different from criminal neonaticide, in which a mother, who has concealed her pregnancy and given birth in secret, kills the infant immediately after parturition. This was a major public health problem in Europe during the nineteenth century. Its frequency has dwindled as a result of contraception, a relaxation of the abortion laws, and changed attitudes to single motherhood; but it still occurs.

The mental state of mothers who kill the newborn can be deduced from the methods used. Suffocation is by far the most common,[66] and this testifies to the mother's panic, faced by a crying baby. In a minority, brutal head injuries, stabbing, or decapitation testify to rage and hatred.

In Europe, starting in Russia in 1647, the public has gradually taken a humane view of this felony. By 1881, all European states, with the exception of England and Wales, made a distinction between infanticide and other forms of murder, and assigned a more lenient penalty. England and Wales at last came into line with the Infanticide Act 1922. In some American states no distinction is made between this and other forms of murder.

There has been much debate whether the defence of insanity can be invoked. Most of these babies die when the mother is in the grip of an emotional crisis—seized by fear or fury. This is not generally acceptable in law as evidence of insanity, which is defined as a defect of reason. However, impairment of consciousness undoubtedly occurs during labour (see above); it is rare in hospital practice, but may be more common in clandestine deliveries, and is hard to exclude. If the defence is burdened with the proof of insanity, there can be no valid evidence in unwitnessed deliveries; but there is the possibility of a miscarriage of justice—that a mother, who killed her baby when her consciousness was clouded, is wrongly condemned.

Filicide

The majority of murdered children are killed by their parents, and the majority of female murderers kill their own child. A survey in Queensland gave an estimate of the frequency: of 49 infanticides between 1969 and 1978, there were 11 neonaticides and 38 filicides; this is about 3 in 100 000 per year of children under 5 years of age.[67] In Sweden, between 1971 and 1980, there were 79 cases involving 96 children—an annual rate of 2 in 100 000 children under the age of 5 years.[68] There are a variety of causes.

- **Depression** This is the most common cause. Studies of convicted mothers underestimate the frequency of depressive filicide, because many complete suicide. Melancholic filicide is committed in the belief that the child's best interests are being served (delusional mercy-killing). Mothers surviving depressive filicide usually make no attempt to conceal the crime; they confess and seek punishment. Mothers may kill more than one child, but family murder seems more common in men.
- **Child abuse** This is the other relatively common cause. Death results from ill-tempered assaults or overzealous punishment, without homicidal intent. Fathers are often involved.
- **Psychosis** In non-affective psychosis, filicide may occur if delusions involve the child, or as a result of command hallucinations.
- **Trance states** A few filicides have occurred during epileptic automatism or somnambulism.

Not all parental child killing occurs as a complication of mental illness. Unwanted infants are occasionally murdered in cold blood. Euthanasia of an incurably ill and suffering child can also occur.

Further information

The only modern texts that cover most of this subject are *Motherhood and Mental Health* (reference 30), and *Psychological Aspects of Women's Health Care* (2002), editors D. E. Stewart & N. L. Stotland, Washington, American Psychiatric Press.

References

1. Hitzig, J.E. (1827). Mord in einem durch Eintreten des Monatsflusses herbeigeführten unfreien Zustande. *Zeitschrift für Criminal-rechts-pflege*, **6**, 237–331.
2. Brière de Boismont, A. (1851). Recherches bibliographiques et cliniques sur la folie puerpérale, précédées d'un aperçu sur les rapports de la menstruation et de l'aliénation mentale. *Annales Médico–psychologiques*, **3**, 574–610.
3. Schlager, L. (1858). Die Bedeutung des Menstrualprocesses und seiner Anomalieen für die Entwicklung und den Verlauf der psychischen Störungen. *Zeitschrift für Psychiatrie*, **15**, 459–98.
4. Bancroft, J. (1993). The premenstrual syndrome—a reappraisal of the concept and the evidence. *Psychological Medicine*, (Suppl. 24), 1–47.
5. von Krafft-Ebing, R. (1902). *Psychosis Menstrualis. Eine klinisch–forensische Studie*. Enke, Stuttgart.
6. Brockington, I.F. (2005). Menstrual psychosis. *World Psychiatry*, **4**, 9–17.
7. Kitayama, I., Yamaguchi, T., Harada, M., *et al.* (1984). Periodic psychoses and hypothalamo-pituitary function. *Mie Medical Journal*, **34**, 127–38.
8. Parker, P.J. (1983). Motivation of surrogate mothers: initial findings. *The American Journal of Psychiatry*, **140**, 117–18.
9. Bivin, G.D. and Klinger, M.P. (1937). *Pseudocyesis*. Principia Press, Bloomington.
10. Ekblad, M. (1963). Social-psychiatric prognosis after sterilisation of women without children. *Acta Psychiatrica Scandinavica*, **39**, 481–514.
11. Cooper, P., Gath, D., Rose, N., *et al.* (1982). Psychological sequelae to elective sterilisation: a prospective study. *British Medical Journal*, **284**, 461–4.
12. Bledin, K.D., Cooper, J.E., MacKenzie, S., *et al.* (1984). Psychological sequelae of female sterilisation: short-term outcome in a prospective controlled study. *Psychological Medicine*, **14**, 379–90.

13. Gath, D., Rose, N., Bond, A., *et al*. (1995). Hysterectomy and psychiatric disorder: are the levels of psychiatric morbidity falling? *Psychological Medicine*, **25**, 277–83.
14. Martin, R.L., Roberts, W.V., and Clayton, P.J. (1980). Psychiatric status after hysterectomy. A one-year prospective follow-up. *The Journal of the American Medical Association*, **244**, 350–3.
15. Aziz, A., Bergquist, C., Nordholm, L., *et al.* (2005). Prophylactic oöphorectomy at elective hysterectomy. Effects on psychological well-being at 1-year follow-up and its correlations to sexuality. *Maturitas*, **16**, 349–57.
16. Cohen, R.L. (1988). *Psychiatric consultation in childbirth settings*. Plenum, New York.
17. Cartwright, A. (1988). Unintended pregnancies that lead to babies. *Social Science and Medicine*, **27**, 249–54.
18. Brockington, I.F. (2006). *Eileithyia's mischief: the organic psychoses of pregnancy, parturition and the puerperium*. Eyry Press, Bredenbury.
19. Wessel, J. and Buscher, U. (2002). Denial of pregnancy: a population based study. *British Medical Journal*, **324**, 458.
20. Nirmal, D., Thijs, I., Bethel, J., *et al.* (2006). The incidence and outcome of concealed pregnancies among hospital deliveries: an 11-year population-based study in South Glamorgan. *The Journal of Obstetrics and Gynaecology Research*, **26**, 118–21.
21. Condon, J.T. (1987). The battered foetus syndrome. *The Journal of Nervous and Mental Disease*, **175**, 722–5.
22. Weir, J.G. (1984). Suicide during pregnancy in London 1943–1962. In *Suicide in pregnancy* (eds. G.J. Kleiner and W.M. Greston), pp. 59–62. Wright, Boston.
23. Ulleland, C.N. (1972). The offspring of alcoholic mothers. *Annals of the New York Academy of Sciences*, **197**, 167–9.
24. Lemoine, P., Harousseau, H., Borteyru, J.P., *et al.* (1968). Les enfants de parents alcooliques: anomalies observées. *Ouest-Médical*, **25**, 476–82.
25. Brinch, M., Isager, T., and Tolstrup, K. (1988). Anorexia nervosa and motherhood: reproduction pattern and mothering behaviour of 50 women. *Acta Psychiatrica Scandinavica*, **77**, 611–17.
26. Morgan, J.F., Lacey, J.H., and Sedgwick, P.M. (1999). Impact of pregnancy on bulimia nervosa. *The British Journal of Psychiatry*, **174**, 135–40.
27. Jureidini, J. (1993). Obstetric factitious disorder and munchausen syndrome by proxy. *The Journal of Nervous and Mental Disease*, **181**, 135–7.
28. Kendell, R.E., Chalmers, J.C., and Platz, C. (1987). Epidemiology of puerperal psychoses. *The British Journal of Psychiatry*, **150**, 662–73.
29. Grof, P., Robbins, W., Alda, M., *et al*. (2000). Protective effect of pregnancy in women with lithium-responsive bipolar disorder. *Journal of Affective Disorders*, **61**, 31–9.
30. Brockington, I.F. (1996). *Motherhood and mental health*, pp. 92, 111–12. Oxford University Press, Oxford.
31. Engelhard, J.L.B. (1912). Über Generationspsychosen und der Einflus des Gestationsperiode auf schon bestehende psychische und neurologische Krankheiten. *Zeitschrift für Geburtshülfe und Gynäkologie*, **70**, 727–812.
32. Gissler, M., Hemminki, E., and Lönnqvist, J. (1996). Suicides after pregnancy in Finland, 1987–1994: register linkage study. *British Medical Journal*, **313**, 1431–4.
33. David, H.P. (1985). Post-abortion and post-partum psychiatric hospitalization. *CIBA Foundation Symposium*, **115**, 150–64.
34. Nagy, M. (1948). The child's theories concerning death. *The Journal of Genetic Psychology*, **73**, 3–27.
35. Smialek, Z. (1978). Observations on immediate reactions of families to sudden infant death. *Pediatrics*, **62**, 160–5.
36. Hughes, P., Turton, P., Hopper, E., *et al.* (2002). Assessment of guidelines for good practice in psychosocial care of mothers after stillbirth: a cohort study. *Lancet*, **360**, 114–48.
37. Condon, J. T. (1986). Psychological disability in women who relinquish a baby for adoption. *The Medical Journal of Australia*, **144**, 117–19.
38. Brockington, I.F., Macdonald, E., and Wainscott, G. (2006). Anxiety, obsessions and morbid preoccupations in pregnancy and the puerperium. *Archives of Women's Mental Health*, **9**, 253–64.
39. Bydlowski, M. and Raoul-Duval, A. (1978). Un avatar psychique méconnu de la puerpéralité: la nevrose traumatique post-obstétricale. *Perspectives Psychiatriques*, **4**, 321–8.
40. Wenzel, A., Haugen, E.N., Jackson, L.C., *et al.* (2005). Anxiety symptoms and disorders at eight weeks postpartum. *Journal of Anxiety Disorders*, **19**, 295–311.
41. Hertzberg, T. and Wahlbeck, K. (1999). The impact of pregnancy and puerperium on panic disorder: a review. *Journal of Psychosomatic Obstetrics and Gynaecology*, **20**, 59–64.
42. De Armond, M. (1954). A type of post partum anxiety reaction. *Diseases of the Nervous System*, **15**, 26–9.
43. Sved-Williams, A.E. (1992). Phobic reactions of mothers to their own babies. *The Australian and New Zealand Journal of Psychiatry*, **26**, 631–8.
44. Weightman, J., Dalal, B.M., and Brockington, I.F. (1999). Pathological fear of cot death. *Psychopathology*, **167**, 246–9.
45. Maina, G., Vaschetto, P., Ziero, S., *et al.* (2001). Il postpartum come fattore di rischio specifico per l'esordio del disturbo ossessivo-compulsivo: studio clinico controllato. *Epidemiologia e Psychiatria Sociale*, **10**, 90–5.
46. Labad, J., Menchon, J.M., Alonso, P., *et al.* (2005). Female reproductive cycle and obsessive-compulsive disorder. *The Journal of Clinical Psychiatry*, **66**, 428–35.
47. Jennings, K.D., Ross, S., Popper, S., *et al.* (1999). Thoughts of harming infants in depressed and non-depressed mothers. *Journal of Affective Disorders*, **54**, 21–8.
48. Gordon, R.E. and Gordon, K.K. (1959). Social factors in the prediction and treatment of emotional disorders of pregnancy. *American Journal of Obstetrics and Gynecology*, **77**, 1074–83.
49. Affonso, D.D., De, A.K., Horowitz, J.A., *et al.* (2000). An international study exploring levels of postpartum depressive symptomatology. *Journal of Psychosomatic Research*, **49**, 207–16.
50. Malmquist, A. and Kaij, L. (1971). Motherhood and childlessness in monozygotic twins. Part II. The influence of motherhood on health. *The British Journal of Psychiatry*, **118**, 22–8.
51. Høyer, G. and Lund, E. (1993). Suicide among women related to number of children in marriage. *Archives of General Psychiatry*, **50**, 134–7.
52. Pound, A., Puckering, C., Cox, A., *et al.* (1988). The impact of maternal depression on young children. *British Journal of Psychotherapy*, **4**, 240–52.
53. Cox, J.L., Holden, J.M., and Sagovsky, R. (1987). Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *The British Journal of Psychiatry*, **150**, 782–6.
54. Epperson, C.N., Czarkowski, K.A., Ward-O'Brien, D., *et al.* (2001). Maternal sertraline treatment and serotonin transport in breast-feeding mother-infant pairs. *The American Journal of Psychiatry*, **158**, 1631–7.
55. Brockington, I.F., Aucamp, H.M., and Fraser, C. (2006). Severe disorders of the mother-infant relationship: definitions and frequency. *Archives of Women's Mental Health*, **9**, 243–52.
56. Righetti-Veltema, M., Conne-Perréard, E., Bousquet, A., *et al.* (2002). Postpartum depression and mother-infant relationship at 3 months. *Journal of Affective Disorders*, **70**, 291–306.
57. Bernazzani, O., Marks, M.N., Bifulco, A., *et al.* (2005). Assessing psychosocial risk in pregnant/postpartum women using the contextual assessment of maternity experience (CAME). *Social Psychiatry and Psychiatric Epidemiology*, **40**, 497–508.
58. Murray, L., Hipwell, A., and Hooper, R. (1996). The cognitive development of 5-year-old children of postnatally depressed mothers. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **37**, 927–35.

59. Ndosi, N.K. and Mtawali, M.L. (2002). The nature of puerperal psychosis at Muhimbili national hospital: its co-morbidity, and associated main obstetric and social factors. *African Journal of Reproductive Health*, **6**, 41–9.
60. Srinavasan, K. (1983). Cerebral venous and arterial thrombosis in pregnancy and the puerperium: a study of 135 patients. *Angiology*, **34**, 731–46.
61. Osiander, F.B. (1797). *Neue Denkwürdigkeiten für Aerzte und Geburtshelfer*, pp. 52–128. Rosenbusch, Göttingen.
62. Esquirol, J.E.D. (1818). Observations sur l'aliénation mentale à la suite de couches. *Journal Général de Médecine, de Chirurgie et de Pharmacie Françaises et Étrangères* (Series 2), **1**, 148–64.
63. Terp, I.M. and Mortensen, P.B. (1998). Post-partum psychoses: clinical diagnoses and relative risk of admission after parturition. *The British Journal of Psychiatry*, **172**, 521–6.
64. Jones, I. and Craddock, N. (2001). Familiarity of the puerperal trigger in bipolar disorder: results of a family study. *The American Journal of Psychiatry*, **158**, 913–17.
65. Main, T.F. (1958). Mothers with children in a psychiatric hospital. *Lancet*, **ii**, 845–7.
66. Tardieu, A. (1868). *Étude Médico–Légale sur l'Infanticide*. Baillière, Paris.
67. Wilkey, I., Pearn, J., Petrie, G., *et al*. (1982). Neonaticide, infanticide and child homicide. *Medicine, Science and the Law*, **22**, 31–4.
68. Somander, L.K.H. and Rammer, L.M. (1991). Intra- and extrafamilial child homicide in Sweden 1971–1980. *Child Abuse & Neglect*, **15**, 45–55.

5.5

Management of psychiatric disorders in medically ill patients, including emergencies

Pier Maria Furlan and Luca Ostacoli

The coexistence of psychiatric disorders in patients with medical illnesses may influence both the diagnosis and the course of the illness by their effects on pathophysiological, diagnostic, and therapeutic processes. There may also be effects on patients' collaboration with treatment and on their relationships with health care staff. Several factors change the management of, medical illnesses and psychiatric disorders, and their inter-relation

- increased life-expectancy and increasing survival of people with-severe illness alter the risk of other medical and psychiatric disorders;
- social changes affecting family structure can affect care giving. Other social factors include changes in the role of women (work, delayed maternity); increased immigration with consequent cultural diversity including different concepts of medical and psychiatric disorders (see Chapter 1.3.2);
- increased use of medication in medical and in psychiatric treatment, and changes in the organization of health care and social assistance from hospital-based to community-based.

This chapter describes how to recognize, treat and manage psychiatric disorders in medical illnesses.

The frequency of psychiatric disorders among the medically ill

The prevalence of psychiatric disorders in medical illnesses ranges from 16–60 per cent depending on the research methodology (self-reports or interviews; inclusion or exclusion of somatic items), the setting (out-patient or hospitalized), and the sample. In general, the frequency of psychiatric disorders in patients with heart disease[1] (coronary disease, heart failure), gastrointestinal diseases (irritable bowel syndrome), lung diseases (asthma, chronic bronchitis), and diabetes is 15–20 per cent. In patients with cancer and chronic pain it is 30–40 per cent; and in neurological diseases (Parkinson's, multiple sclerosis, epilepsy) and dialysis it is 50 per cent. Ten to 20 per cent of patients have sub-threshold symptoms that nevertheless influence psychosocial functioning. The prevalence of psychiatric disorders among family members of people with chronic disabling conditions is only slightly lower. The most frequent are organic mental disorders (5–44 per cent), followed by substance abuse (10–25 per cent), anxiety disorders (10–30 per cent), mood disorders (9–13 per cent), personality disorders (6–9 per cent), somatoform disorders (5–9 per cent), mania and psychosis (1 per cent). Recognition by medical doctors is below 50 per cent and the referral rate to liaison services is approximately 1–3 per cent.

The frequency of medical illnesses among psychiatric patients

The most severe psychiatric disorders are frequently associated with social isolation, difficult relations with health-care providers, poor adherence to treatment, unhealthy lifestyle[2] (nutrition, smoking, hygiene), side-effects of medication and substance dependence. The presence of psychiatric symptoms can also lead to failure to recognize physical symptoms. And yet some medical illnesses are more frequent in people with schizophrenia than in the general population.[3] These conditions are cardiovascular risk (9.4 per cent in men, 7 per cent in women), diabetes (13 per cent), hypertension (27 per cent) and chronic conditions in general (41 per cent).[4]

Table 5.5.1 Prevalence of medical illnesses in patients with mood disorders

Disease	%
Hypertension	18.1–34.8
Stroke	1.7–1.9
Headache	19.3
Chronic pulmonary disease	10.6–12.9
Hypothyroidism	9.6
Obesity	4.6
Alcohol abuse	12.2–24.7
Nicotine	9.1–12.6
Illicit drug abuse	9.7

The diagnosis of psychiatric disorder in medically-ill patients

Anxiety, fear, demoralization, a sense of loss, decreased pleasure, and thoughts of death are frequent in advanced debilitating physical disease even when there is no coexisting anxiety or depressive disorder.

Physical disease and its treatment may cause somatic symptoms similar to those of psychiatric disorders. And the 'aetiological' criterion of the DSM-IV-TR that requires exclusion of a physical cause is often difficult to apply in advanced medical illness, as are the criteria for depression. Endicott[5] proposed replacing the four somatic items for depression (fatigue, insomnia, weight-loss, and difficulty in concentrating) with four psychological symptoms: depressed appearance, social withdrawal, brooding, non-reactive mood. However, this proposal risks excluding somatic symptoms which are a core manifestation of more severe forms of depressive disorder. In doubtful cases, an inclusive approach to somatic symptoms is preferable, and the risk of severe psychiatric disorders should not be underestimated.

Self-abasement and guilt are less frequent in medically ill patients. In assessing guilt, ethnic and cultural factors must be taken into account, for example, feelings of guilt are uncommon in depressed Arabs, whereas somatization is common.

Some syndromes in medically ill patients do not correspond to standard diagnostic categories but nevertheless influence functioning and the course of disease. The Diagnostic Criteria for Psychosomatic Research[6] mention illness denial, thanatophobia, demoralization, and alexithymia. In medical illness, psychiatric disorders may manifest with somatic symptoms (see Chapter 5.2.3).

Table 5.5.2 gives some broad indications for the differential diagnosis of psychiatric disorders in the presence of medical illness.[7]

To be classed as a psychological reaction, the psychiatric disorder must develop at the same time as the onset of the medical illness or the treatment. In some conditions such as, pancreatic cancer, multiple sclerosis, the onset of psychiatric disorder may precede the recognition of the medical illness (e.g. Multidimensional evaluation of the care requirements is essential and codified approaches exist).[8]

Atypical symptoms occur in psychiatric disorders due to medical conditions. Drium is often complex with auditory hallucinations prevailing, whereas tactile, olfactory, and gustatory hallucinations are rare.

Causes of psychiatric illness among medical patients

These are both psychological (see Chapter 5.6). and medical (see Chapter 5.3. 4.

Course and prognosis

If properly treated, psychiatric disorders in medically ill patients have the same prognosis as those occurring without medical illness, except in some very advanced and debilitating cases, and in these, the few reported studies give contrasting results. Psychiatric disorders may significantly influence the outcome of the medical condition. Depression is associated with an increased risk for subsequent development of ischaemic heart disease, Parkinson's disease, Alzheimer's disease (and other dementias) and medical diseases in general. It is an independent predictor of severe complications in diabetes and of mortality in ischaemic heart disease, heart failure,[9] stroke, dementia, cancer and HIV. Anxiety may exacerbate angina, arrhythmia, asthma, movement disorders, hypertension and irritable bowel syndrome and is associated with increased health-seeking behaviour and prescription of inappropriate drugs.

Delirium is reversible in 70–80 per cent of cases, but in terminally ill patients may be progressive and intractable, and is associated with increased short-term mortality.[10] Mania and psychosis may worsen the medical outcome due to behavioural alterations, poor adherence and increased drug adverse effects.

Table 5.5.2 Differential diagnosis among psychiatric disorders (PD) in medical illnesses (MI)

	Comorbidity (MI – PD)	Latent (PD)	Psychological Reaction	Psychoorganic PD
Onset with mi	-	+	+	+/-
Medical aetiology	-	-/+	-	+
Life events	-	+/-	+	-
Personal/family medical history	+	+	-	-
Cognitive disorders	-	-	-	+
Altered awareness	-	-	-	+
Fluctuation in severity of psychic symptoms	-	-	+	+
Atypical psychic symptoms	-	-	-	+
Self-abasement	+	+/-	-	-
Family history for psychic symptoms	+	+/-	-	-
Empathy of doctor	-/+	+/-	+	-
Response to psychiatric treatment	+/-	+	+	-

Treatment

In specific populations of patients such as those with diabetes, asthma, myocardial infarction, irritable bowel syndrome, cancer, Parkinson's disease, multiple sclerosis, and rheumatic diseases, psychosocial interventions and a variety of psychological treatments have a positive effect on psychiatric disorders, and the quality of life and relationships.[11] In medical conditions requiring active patient participation, psychological treatments improve adherence to the therapeutic programmes. In diabetes they reduce glycosylated haemoglobin, in Parkinson's they produce cognitive and motor improvement. Such treatments can reduce physical symptoms including pain, nausea, dyspnoea and disability. For other indices such as mortality in heart attack, longevity in cancer, severity of hypertension and peptic ulcer, and inflammatory activity in rheumatoid arthritis, there are psychological benefits and improvements in quality of life but poor effects on medical outcome. Indeed, in the more severe psychiatric disorders, combined treatment with psychopharmacological drugs is more effective. Studies of cost-effectiveness and length of hospitalization have reported conflicting results.[12]

Management

The consultation process

(a) Forming an alliance with the medical team

The psychiatrist should initially aim to work with medical staff on their clinical rounds, and interview patients who have no psychiatric disorder to learn what it means to have the medical condition and to undergo its treatment. The psychiatrist should be present at informal discussions, such as those during coffee breaks,and share the clinical team's emotional experiences.

(b) Interview with medical team

The aim is to evaluate the clinical situation, review the medical records; identify the most significant reason for referral and why the consultation has been made at this time; identify the team's approach to the patient, clarify whether the patient has been informed of the consultation and in what way. If the referral seems inappropriate for the patient, does this reflect a problem within the medical team such as burn out?

(c) Interview with the patient

The interview should move dynamically between the 'objective' position (clinical data, psychosocial information) and the 'subjective' position. Feelings evoked in the psychiatrist frequently mirror perceptions of the patient and the medical team. To recognize them helps empathetic relations and emotional containment. A protocol for the interview is shown below. [13]

(d) Reporting back to the medical team

Reporting should be clear and concise with both a verbal and a written report. The focus should be on the reason for referral. Risk factors and points of strength should be identified. A verbal report of an 'image' such as an episode fom the patient's experience, a memory or even a dream can sometimes aid empathetic understanding by the medical team. Practical advice on management should be provided.

(e) Psychopharmacological treatment

The psychopharmacological treatment of medically ill patients may be difficult for several reasons including the stigma associated with psychiatric disorders, weariness with the many medical treatments already undergone, increased sensitivity to side-effects due to pharmacokinetic changes produced by interactions with medical drugs and any underlying liver or renal disease.[14] However when adherence is adequate, the therapeutic response is similar to that of patients without medical illnesses. Nevertheless, the prescription of medication must not replace receptiveness to the patients problems and emotional support, which are frequently the most effective intervention.

It is often useful to offer the patient a drug not for symptoms such as depression when these arise in a discouraging physical situation because this may be seen as disparaging—but for other

Goal	Approach
Overcome stigma	The medical team: present the psychiatrist to the patient as a team member and motivate consultation The psychiatrist: explores any ambivalence or negative feelings empathetically (it may not be clear to the patient why a psychiatrist is discussing his/her disorder)
Open questions: background and information	Why is the patient hospitalized? What does he/she think of the illness?
Principal emotions and fears	What does the patient feel? What is he/she most afraid of?
Consider constructive ways of coping; and discuss dysfunction	What and who most help the patient to overcome difficulties; who is important, what other resources are available
Completing the information	Medical history and current quality of life
Develop a shared understanding of the situation in which medical and psychological aspects are linked and not viewed as alternatives.	"Normalize" emotional disorders as reactions to illness that may amplify symptoms and influence course. Describe physical mechanisms of symptom production such as muscular contraction, vasodilation/constriction, hyperventilation, asthenia, inactivity, immune defences. Use clear, descriptive language with imagery: e.g. tension is like a tight shoe.
Defining goals and reducing unrealistic expectations	Discuss the main problems Aim to improve problems, not solve them. Set realistic goals
Propose a treatment plan	What interventions, "first steps" Practical advice about day-to-day matters such as interpersonal relations, how to reduce tension

realistic goals. The specific contribution that the drug can make to achieving these goals should be explained and a clear description provided of its somatic effects. The psychiatrist should also:

- Aim to simplify treatment as far as possible. The distinction between 'psychiatric' and 'medical' drugs is arbitrary since many drugs have both physical and psychological effects.
- Rationalize treatment using the fewest possible drugs. First choice should be drugs that act on more than one symptom, psychological or medical (e.g. reduce agitation and nausea, insomnia and hyporexia, depression and headache, pain and anxiety).
- Consider side-effects and interactions before deciding treatment including their time of onset. Thus, SSRI should be prescribed at least one week apart from any treatment with significant gastrointestinal side-effects. Investigate complementary drug use: e.g. 20–40 per cent of oncology patients take herbal remedies, many of which have significant pharmacological interactions (see Chapter 6.2.9).
- Prefer drugs that were effective for the patient in the past; starting with low doses and increasing them gradually.
- Sometimes the best action is to discontinue a drug.

Psychosocial treatments

Differences from traditional psychotherapy

The major aims are to control distress, maintain self-respect, maintain significant relations and doctor-patient communication, work through information, and develop adaptive coping mechanisms.

Timing: At times of crisis (e.g. immediately after the communication of a serious diagnosis), the need for control is paramount. At the onset of complex diseases, short cycles of interviews may prevent subsequent psychiatric disorder.

Flexibility: This is needed due to variations over time in the conditions of treatment (in hospital, or as a day patient or outpatient), the symptoms, and the motivation of the patient, who may alternate between the need for emotional sharing and moments of self-withdrawal. Existential uncertainty is reflected in relations with the psychiatrist and each interview is an entity in itself.

Eclecticism: The complexity of the situation often requires integration of different approaches at different times: expressive, cognitive, body-mediated or psycho educational. The intervention may focus on the patient, on family members or on the medical team. Flexibility and eclecticism must originate from the integration of the patient's needs and the psychiatrist's empathic and comprehensive evaluation of these needs.

Regression: Illness, hospitalization, fear, and the 'invalid role' may make a patient, who would otherwise refuse it, become receptive to the psychiatrist's intervention. During medical illness, emotional defences are more fluid, the relationship with the psychiatrist may form more rapidly and short interventions can be effective. After discharge the psychiatrist must be ready to change approach.

Existential context: A severe disease often casts doubt on the meaning of existence. Patients may fall into despair or ask themselves what really matters. They may want to reorganize their lives around new priorities. In these circumstances, interviews should focus on the present, abandoning the 'past-future' approach, and define the psychiatric disorder as an accentuation of natural human emotions. Fear should be dealt with directly, including fear of dying, and the patient's relatives, including their doctors, should be helped to provide support since fear is greater when experienced alone.

Physical contact: Faced with severe disease, some patients 'speak' only after having been touched. Sometimes bodily contact, for example through simple massage, may keep open communication with family members and staff, even without the use of words. A glance, a voice, the sense of touch, or other bodily sensations may form an intense dialogue between patient and therapist and be the most effective way to understand how aware the patient is of his/her condition.

Treatment of urgent situations

Prevention

- Medical treatment should address the overall quality of life, including the gap between expectations and reality. Many medical interviews target practical matters but place less emphasis on working through expectations. A doctor-patient relationship capable of offering relief while gradually reducing the gap between hope and reality is fundamental.
- The patient's chief supports are family members and caregivers. Improving their psychological skills through simple psychoeducational programmes may be more cost-effective than generally increasing psychiatric consultations.
- Psychosocial and biological risk factors should be identified early, including both current factors and those in the medical history.

Acute anxiety

Treatment of acute anxiety should maximize the beneficial effects of the doctor-patient relationship as well as providing specific psychosocial and pharmacological interventions. At times of crisis, one-on-one companionship is useful, sometimes with simple relaxation or massage that family members can provide. When a serious medical diagnosis is communicated, it is helpful to listen, and for medical staff to be receptive to emotional reactions in the subsequent 24 hours; solitude should be alleviated, if necessary with the help of voluntary workers. Benzodiazepines should be prescribed only when really necessary and for short periods to reduce the risk of tolerance and addiction. Benzodiazepines may reduce respiratory function further in patients who retain CO_2, thereby worsening their asthenia. If prolonged pharmacological therapy is necessary, sedative antidepressants may be an alternative to benzodiazepines.

Panic attacks: Somatic symptoms of a panic attack may be confused with an excacerbation of medical conditions such as chest pain, irritable bowel syndrome, and asthma. Treatment with a selective serotonin reuptake inhibitor is usually effective but half of these patients require long-term treatment since relapse is common after discontinuation. Psychological treatment[15] (see Chapter 4.7.3) can be effective.

Post-traumatic stress disorder: Stress factors in medical contexts may cause and prolong this disorder. Such factors include acute medical events, intensive care, post-confusional reactions, communication of a serious diagnosis, and presence at unexpected deaths. Pharmacological therapy has limited results but psychological treatments may be effective (see Chapter 4.6.2), Risk prevention is important.

Depression

Treatment should be based on prevention, and psycho education, psychosocial and pharmacological interventions.

Antidepressants: The choice of drug should be based on the side-effect profile and interactions (see Chapter 6.2.3). The somatic effect of treating depression can be important. Cognitive and motor functions can improve in stroke patients, glycaemic control can improve in diabetics, as can chronic pain, and dyspnoea in lung-disease. SSRIs may produce gastrointestinal side effects, bleeding due to platelet dysfunction, so that monitoring is essential in patients on anticoagulants. Venlafaxine in high doses may increase blood pressure. Mirtazapine has no sexual side-effects, has anti-nausea activity and stimulates the appetite, but may cause weight gain and sedation. TCAs, being anticholinergic, affecting heart conduction and the peripheral autonomic nervous system, are contraindicated in heart disease, cognitive impairment, orthostatic hypotension, hypertrophic prostate, glaucoma and epilepsy. Secondary amines, such as nortriptyline and desipramine, are preferable.

Methylphenidate may be useful in particular medical situations, such as palliative care or advanced cancer, to alleviate fatigue, increase appetite, and reduce opiate-induced sedation. It may elevate mood and has rapid onset. Agitation and insomnia are the main problems.

Personality disorders

Personality disorders, particularly antisocial and borderline disorders, may create emergencies due either to aggressive behaviour towards self or others or to conflict within the team. Screening for substance abuse and depression is necessary since these increase impulsiveness. Patients can become aggressive when their demands are not met, e.g. for increased painkillers, or interviews with their doctors. Staff require support to manage conflicts and frustration, and to maintain a caring approach.[16]

Staff must be empathetic but at the same time act as a team. Limits and rules should be established at admission treatment to prevent escalation of aggression. Interviews must be in conditions of safety; if necessary with other staff or even police present. The agreed plan must be respected by all staff, avoiding the extremes of excessive permissiveness and excessive rigidity, which may be induced by the patient's disorder. Patients with personality disorders often use truth-based observations pathologically to exploit the therapist's and institution's weak points. Recognizing what is true in their criticisms may improve patient management.

Delirium and dementia (see Chapters 4.1.1 and 4.1.13)

Mania

A manic crisis affects the care of the medical condition dramatically both through excessive activity, weight-loss, reduction of sleep, and through non-compliance with treatment and conflict with staff. The medical history should be evaluated for disease-related stress factors which may trigger crises. Manic symptoms recognized early together with possible causes of insomnia such as pain, nocturia, dyspnoea, or environmental factors including noise, unsuitable temperature, frequent entrance of night-staff for other patients. Where possible, precipitating factors should be removed. When this is impossible (e.g. a requirement for high-dose cortisone), treatment is as that of mania in other situations (see Chapter 4.5.8) taking medication side-effects into account (see Chapter 6.2.4). Antipsychotic drugs are generally faster acting at lower doses than in primary mania. More potent antipsychotics are preferable since they have lower anticholinergic and alpha-blocking effects. It is useful to focus on symptoms the patient finds disturbing rather than ego-syntonic symptoms. Restraint or hospitalization in a psychiatric ward is necessary when hyperactivity could compromise physical safety or the underlying condition.

Psychosis

Close collaboration between the psychiatrist and the medical team is necessary, with the most appropriate environment for treatment, medical or psychiatric unit, being decided case-by-case. Medical investigations and interviews should be simplified and clear basic information provided. Continuity of care among medical professionals is important. The psychological effect of medical treatment should be assessed. When there are hallucinations, medical staff should be educated about their nature and supported, so that they do not criticize any content expressed by the patient, but are able to empathize with the patient's distress. Management may require constant observation and in some cases restraint and involuntary treatment. Medically ill patients are especially sensitive to the side-effects of antipsychotic drugs but generally respond to lower doses. The choice of drug and dose are based on the side-effects and the underlying disease (see Chapter 6.2.2)). Cardiac pathology and QTc enlargement require ECG monitoring and ziprasidone, haloperidol, chlorpromazine, and thioridazine are contraindicated. Caution is also necessary in alcoholics, and patients with hypokalemia and hypomagnesemia.[17]

Extra-pyramidal effects of high doses of typical neuroleptics may cause laryngospasm and affect the diaphragm, worsening any respiratory insufficiency. For chronic treatment, newer drugs are preferable. Caution is required in hepatic and renal diseases and where there is a risk of epileptic fits. Fluid intake should be monitored carefully because dehydration increases the risk of neuroleptic malignant syndrome. In secondary psychosis, the family should be informed and supported and, after the psychotic episode has been resolved, the patient must be helped to work through the disorientating experience.

Aggression and restraint

Aggression should be assessed, thoroughly defining the sequence of events that preceded it and the role of medical and psychiatric factors, stressful events and environmental factors. Major predictors related to the patient include a history of violence, difficulty in communication, and the psychological condition. Predictors related to the environment include: crowding, and a medical team unreceptive to the patient's discomfort. The best strategy is prevention through providing information in a clear and empathetic way, recognizing risk situations and early signs of agitation, and using de-escalation techniques if necessary with pharmacological support (see also Chapter 11.14).

Environmental measures to limit psychomotor agitation include: making the patient as comfortable as possible, reducing background noise, removing potentially harmful objects, and identifying situations of which the patients is intolerant. Patients can be

distracted by walking or occupational activities, avoiding a forced stay in bed and making the timetable flexible. If possible, family members can be asked to collaboration.

Where permitted under national law, restraint may become necessary during unmanageable agitation and aggression in the course of an acute psychiatric disorder, in confusional states, especially in the elderly and aftersurgery, with metabolic diseases or during abstinence from alcohol and drugs (seen especially during the first few days of hospitalization).

If restraint is permissible, staff training is necessary because improper use may cause injury to patients or staff. Restraint should be used only when strictly necessary and for the shortest possible time, with constant monitoring of its continuing necessity and the patient's condition.

In medically ill patients restraint may increase the risk of thromboembolism, reduce respiratory function with increased risk of infection, increase cognitive deficit, pressure ulcers, urinary and intestinal incontinence. When restraint is removed there is an increased risk of falls.

Deliberate self-harm and suicidality (see also Chapter 4.15.4)

In advanced medical illness, suicidal ideation is frequent as an expression of the wish for release from suffering, to regain control over a condition that is perceived as unstoppable, and is wearing the patient down. The actual risk of suicide is slightly above that of the general population. Specific risk factors are: chronic disease, uncontrolled physical symptoms such as pain or dyspnoea, disfiguring surgery, unrecognized psychiatric disorder (delirium, depression, personality disorder), substance abuse, and lack of social support. The most frequent form of self-harm is drug overdose, particularly of analgesics and psychopharmacological drugs. Most lethal suicidal acts are of the impulsive type, in particular jumping. [18] The doctor-patient relationship is often the best help against loss of hope. Safety of the environment is crucial since medical contexts facilitate access to potentially lethal objects and to the possibility of jumping. In cancer, the risk is higher in the first year, decreasing with time after diagnosis. In the end stage of illness e.g. in renal failure, treatment withdrawal may have to be considered, with the: clinical and ethical problems of distinguishing a rational decision by the patient from one affected by a psychiatric disorder or another potentially modifiable factor.

Patients refusing treatment (see also Chapter 1.5.1)

Background culture should be taken into account as some 'overall rejections' are actually related to this rather than to factors in the individual. An alliance with family members is necessary. Factors in the doctor-patient relationship should be evaluated, together with the information that the patient has received. In some cases refusal relates to fear, anger or despair. If it seems that the patient would feel belittled by a psychiatric consultation it is better to support medical staff who are receptive to the patient's emotional sate. This is done patients often then accept psychiatric support more readily. In complex cases, the history of the patient's attitude to treatment should be evaluated on admission to treatment. Aa history of interrupted treatment, changes of doctor or diffidence towards caregivers, may indicate a need for greater efforts to develop good doctor-patient relationships. When obtaining items of medical history from family members, it is important to identify whom the patient trusts most. In relations with the medical team, the most appropriate figure for dialogue should be identified. In some cases, informal conversations, for example in the evening, are more useful than formal medical interviews.

Further information

Levenson, J.L., (2005). *Textbook of psychosomatic medicine*. American Psychiatric Publishing, Washington, DC.

Wise, M.G. and Rundell, J.R., (2005). *Clinical manual of Psychosomatic medicine: A Guide to Consultation–Liaison Psychiatry*. American Psychiatric Press, Washington, DC.

Guthrie, E. and Creed, F. (1996). *Seminars in Liaison Psychiatry*. College seminars series. Royal College of Psychiatrists, London.

Web sites

The European Association for Consultation Liaison Psychiatry and Psychosomatics. http://www.eaclpp.org

International Organization for Consultation- Liaison Psychiatry. http://www. med.monash.edu.au/psychmed/ioclp.

The International College of Psychosomatic Medicine. http://www.icpm.org

References

1. Rutledge, T., Reis, V.A., Linke, S.E., *et al.* (2006). Depression in heart failure a meta-analytic review of prevalence, intervention effects, and associations with clinical outcomes. *Journal of American College of Cardiology,* **48**(8), 1527–37.
2. Michael, T., Compton, M.T., Daumit, G.L., *et al.* (2006). Cigarette Smoking and Overweight/Obesity Among Individuals with Serious Mental Illnesses: A Preventive Perspective. *Harvard Review of Psychiatry,* **14**(4), 212–22.
3. Carney, C.P. and Jones, L., (2006). Medical Comorbidity in women and men with bipolar disorders: a population – based controlled study. *Psychosomatic Medicine,* **68**, 684–91.
4. Goff, D.C., Sullivan, L.M., McEvoy, J.P., *et al.* (2005). A comparison of ten-year cardiac risk estimates in schizophrenia patients from the CATIE study and matched controls. *Schizophrenia Research,* **80**, 45– 53.
5. Endicott, J., (1984). Measurement of depression in patients with cancer. *Cancer,* **53**(Suppl 10), 2243–9.
6. Fava, G.A., Freyberger, H.J., Bech, P., *et al.* (1995). Diagnostic criteria for use in psychosomatic research. *Psychotherapy Psychosomatic,* **63**(1), 1–8.
7. Lipowski, Z.J., (1967) Review of consultation psychiatry and psychosomatic medicine, II: clinical aspects. *Psychosomatic Medicine,* **29**, 201–24.
8. Huyse, F.J., Lyons, J.S., Stiefel, F.C., *et al*, (1999). INTERMED: a method to assess health service needs. I. Development and reliability. *General Hospital Psychiatry,* **21**, 39–48.
9. Barth, J., Schumacher, M. and Lingen, C.H. (2004). Depression as a Risk Factor for Mortality in Patients With Coronary Heart Disease: A Meta-analysis, *Psychosomatic Medicine,* **66,** 802–13.
10. Breitbart, W., Gibson, C. and Tremblay, A., (2002). The delirium experience: delirium recall and delirium-related distress in hospitalized patients with cancer, their spouses/caregivers, and their nurses. *Psychosomatis,* **43**, 183–94.
11. Guthrie, E. and Creed, F. (1996). Treatment methods and their effectiveness. In *Seminars in Liaison Psychiatry* (eds. E., Guthrie, and F., Creed), pp. 238-73. College seminars series. Royal College of Psychiatrists, London.

12. Andreoli, P.B., Citero, Vde A. and Mari Jde, J., (2003). A systematic review of studies of the cost-effectiveness of mental health consultation-liaison interventions in general hospitals. *Psychosomatics,* **44**(6), 499–507.
13. Stuart, M.R. and Lieberman, J.A., (2002). *The fifteen minute hour. Practical therapeutic interventions in Primary care*. Saunders, Elsevier (USA). 14 Levenson JL, (2005). Psychopharmacology. In: Levenson JL, ed. *Textbook of psychosomatic medicine*, American Psychiatric Publishing, Washington, DC.
15. Shapiro, F. (2001). *Eye Movement Desensitization and Reprocessing: Basic Principles, Protocols and Procedures*. Guilford Press, New York.
16. Hay, J.L. and PassiK, S.D. (2000). The cancer patient with borderline personality disorder: suggestions for symptom-focused management in the medical setting. *Psychooncology,* **9**(2), 91–100.
17. Gasper, J.J. and Tsai, C. (2006). Community Behavioral Health Sciences, Community Health Network of San Francisco, San Francisco General Hospital, Revised: *Guidelines for the use of atypical antipsychotics in adults*.
18. Druss, B. and Pincus, H. (2000). Suicidal ideation and suicide attempts in general medical illnesses, *Arch Intern Med,* **160**(10), 1522–6.

5.6

Health psychology

John Weinman and Keith J. Petrie

Introduction

Health psychology is concerned with understanding human behaviour in the context of health, illness, and health care. It is the study of the psychological factors, which determine how people stay healthy, why they become ill, and how they respond to illness and health care.

Health psychology has emerged as a separate discipline in the past 30 years and there are many reasons for its rapid development. An important background factor is the major change in the nature of health problems in industrialized societies during the twentieth century. Chronic illnesses such as heart disease and cancer have become the leading causes of death, and behavioural factors such as smoking, diet, and stress are now recognized as playing a major role in the aetiology and progression of these diseases.[1] The provision of health care has grown enormously and there is an increased awareness of good communication as a central ingredient of medical care and of the importance of such factors as patient satisfaction and quality of life as key outcomes in evaluating the efficacy of medical interventions.

Although health psychology has developed over a similar time period to general hospital/liaison psychiatry and shares some common areas of interest, there are some clear differences between these two fields. Liaison psychiatry has a primary focus on hospital patients, particularly those experiencing psychological difficulties in the face of a physical health problem. In contrast, health psychology has a much broader focus on both healthy and ill populations and on the psychological processes that influence their level of health or their degree of adaptation to disease. Whereas health psychology has been mainly concerned with developing explanations based on theory, for health-related[2] and illness-related behaviour,[3] liaison psychiatry has concentrated on the diagnosis and treatment of either unexplained symptoms or psychiatric disorders occurring in people with medical conditions (see the other chapters in Part 5 of this volume).

In this chapter we provide an overview of the main themes and areas in health psychology. Four broad areas of behaviour will be reviewed, namely behavioural factors influencing health, symptom and illness behaviour, health care behaviour, and treatment behaviour. Inevitably such an overview is selective and the interested reader should seek out a more comprehensive introductory text[4,5] or more in-depth accounts of specific areas.[2,3]

Behavioural factors influencing health

A wide range of behavioural factors can influence health. In the following section there is a focus on stress, personality, and the main theories that have been developed to explain the variation in health-related behaviours.

Stress and health

The term 'stress' is usually used to describe situations, in which individuals are faced with demands that exceed their immediate ability to cope. Stressful situations are typically those that are novel, unpredictable, and uncontrollable as well as those involving change or negative events such as a loss. These situations can give rise to adverse psychological and physiological changes which, in turn, may result in disease.[6]

Stress may have indirect effects on health by increasing levels of risk behaviour (e.g. smoking, alcohol consumption), or may have direct effects on specific physiological mechanisms (e.g. increase in blood pressure) as well as affecting the individual's resistance to disease through suppression of the immune system, or by exacerbating or triggering a disease process in an already vulnerable individual.

A range of behavioural and emotional responses are shown by individuals as they attempt to cope with stressful situations and these are accompanied by autonomic, neuroendocrine, and immunological changes. During stressful episodes, releasing factors from the brain cause the pituitary to release ACTH which gives rise to the release of corticosteroids from the cortex of the adrenal glands. In addition to producing a number of well-known changes associated with the mobilization of both short- and longer-term physical resources (e.g. release of adrenaline (epinephrine) or noradrenaline (norepinephrine), release of glucose, activation of endorphins/encephalins, etc.), these steroids can also have effects on the immune system.[7]

The effects of stress on immunity have sparked the development of the new multi-disciplinary field of psychoneuroimmunology which focuses on the links between psychological, endocrine, and immunological processes (see Chapter 2.3.10). A large amount of work in this area has concentrated on the links between stress and immune function, but less work has focused on impaired immunity and the later development of disease. Acute stressors, such as examinations, or more chronic stressors, such as caring for

a dependent elderly relative, have been shown to lead to deleterious immunological changes. Work has also associated stress with a greater susceptibility to viral infection[8] as well as longer healing times for experimental puncture wounds[9] and wounds from surgical operations.[10] A recent meta-analysis of studies of stress and immunity shows substantial evidence for a relationship between stress and impaired immune system effectiveness, particularly for chronic uncontrollable sources of stress.[7]

Personality and health

Although there is no consistent empirical support for the older idea that different diseases are linked with specific personality types, there is evidence from different, more credible sources that personality factors can influence health and play a role in determining illness in other ways.[11]

Probably the best known work in this area concerns the link between the so-called 'type A' personality and coronary heart disease. The *type A personality* was originally characterized by competitiveness, time urgency, hostility, and related behavioural factors, which were associated with a significantly increased risk of coronary heart disease (CHD). However, it is now thought that only certain components (e.g. anger and hostility) of the original type A formulation are 'pathogenic'.[12]

Type A individuals show a greater physiological reactivity (e.g. in blood pressure and heart rate) to environmental demands and may even generate more demands by their style of behaviour. The more frequent elevations in blood pressure and higher levels of hormonal change, characteristic of this behavioural style, may eventually cause adverse physical changes to the heart and blood vessels. Also, type A individuals are more likely to engage in unhealthy behaviours since they drink more alcohol than type B individuals and, if they smoke, they inhale their cigarette smoke for a longer time.

Type A behaviour is probably the most extensively investigated personality factor in current health psychology research, and there have been interventions developed to change the behaviour pattern, with positive health outcomes.[13] More recently the concept of the *type D personality* has been described as another major psychological risk factor for CHD. Type D refers to the tendency to experience negative emotional states and to inhibit the expression of these emotions in social settings. Type D patients with CHD have been found to have a significantly higher risk of further cardiac morbidity in the short- and longer-term.[14]

More generally, patterns of positive or negative emotional responses, associated with personality, can influence various aspects of health.[11,15] Individuals who are high in negative affect (i.e. experience more negative emotions, particularly anxiety) do not seem to be more prone to disease, but they are more likely to notice bodily changes and symptoms and consequently seek medical help more frequently (see Wiebe and Smith[15] for a more detailed account of negative affect and the links between personality and health).

Another aspect of personality which has been shown to be health protective is optimism, which describes a tendency towards positive expectations in life and which enables individuals to cope better with stressors and engage in healthier lifestyles. There is emerging evidence that optimistic individuals not only cope more effectively with illness and other life crises but also show better health outcomes than those with lower levels of optimism.[16]

Lifestyle and health

The effects on health of behaviours such as smoking and high alcohol use are well documented.[1] There is overwhelming evidence that smokers not only are much more likely to die from lung cancer and other cancers but also have much higher rates of cardiovascular disease and chronic respiratory disorders, particularly emphysema and chronic bronchitis. Moreover, the disease risk is dose related in that higher levels of smoking are more strongly associated with all these diseases. With sustained high levels of alcohol use a different but equally unpleasant spectrum of health problems can be seen. Drinking is a major cause of accidents particularly motoring accidents and can cause liver damage as well as having detrimental effects on brain functioning.[1]

For health psychologists, the key questions about health-risk behaviours concern their origin, their maintenance, and their prevention or treatment. There are diverse determinants of these behaviours since they may start as ways of coping with stress, in response to peer pressure, for pleasure or for a number of other reasons. Similarly, they will be maintained by a variety of psychological, social, and biological factors.

There are many other risky behaviours that cannot be discussed in detail in an overview; these include drug abuse, poor diet, and accidents, and the health effects of all these are also well documented.[1] Although health psychology has an important role to play in describing, explaining, and intervening in all risk behaviours, these problems should not be conceptualized exclusively in individual behavioural terms since they often reflect adverse social circumstances or particular cultural contexts.[17]

The same caveats about the influence of social and cultural factors must also be applied to the understanding of health-protective or health-enhancing behaviours. Prospective cohort studies have confirmed that various daily behaviours (e.g. patterns of eating, sleeping, and exercise) can have significant long-term effects on health.[18] For example, there is now a growing body of evidence to indicate that regular exercise has a beneficial effect on both physical and psychological health.[19] Exercise can reduce the incidence of physical health problems in elderly people and facilitate recovery from heart attack. However, there can be significant problems in ensuring that exercise and other health-promoting activities are adhered to. Interventions need to be planned carefully, because it has been shown that it is usually very difficult to make and maintain changes in health-related behaviour. Information provision is rarely sufficient to promote behaviour change since it is also necessary to elicit and modify beliefs (see below) as well as influencing social networks in order to ensure success.

Beliefs and health-related behaviour

Even though health psychologists acknowledge the importance of situational, dispositional, and socio-cultural factors as determinants of health-related behaviour, most current research has a primary focus on the role of beliefs in explaining variance in health-related behaviour. The most widely used explanatory approaches have been described generically as 'social cognition models' (see Conner and Norman[2] for an excellent overview of these models). These models are based on the premise that, when a person is faced with having to make a decision about a particular health behaviour (e.g. attend for a screening test; wear a seat belt, etc.), their decision-making and behaviour can be best understood in terms of their perceptions or beliefs about the health issue and the behaviour in

question. The best known models here are the Health Belief Model, Theory of Reasoned Action/Theory of Planned Behaviour, and Protection-Motivation Theory. Broadly these models locate the strength of certain beliefs or evaluations of the health threat (e.g. 'is it serious? Is it likely to affect me?') and/or the associated health behaviour ('Is it an acceptable or worthwhile thing for me to do?') as the key determinants of an individual's motivation or intention to carry out the behaviour. More recent models incorporate other beliefs, such as self-efficacy, which reflect the individual's belief about their ability to implement or carry out the health-related behaviour.

For habitual and addictive health-related behaviours (e.g. dietary behaviour; substance abuse) there have also been attempts to develop stage-based models, such as the Precaution Adoption model and the Transtheoretical model[2] as ways of describing the stages which people may go through in evaluating the health issue through to thinking about, planning, and maintaining behaviour change. Although these stage models provide a framework for identifying the patient's state of readiness for a health behaviour change intervention as well as an immediate target for an intervention, the evidence for them is weak and there are now a number of serious critiques of their validity and applicability.[20]

Symptoms and illness behaviour

The psychology of physical symptoms

Understanding how symptoms are perceived is critical to explaining variation in illness behaviour. Psychological factors play an important role in the appraisal of symptoms. There is considerable evidence that bodily symptoms and functions are not perceived with a high degree of accuracy and individuals vary widely in what symptoms are noticed and whether medical help is sought for symptoms.[21]

The probability that individuals will attend to somatic information will depend on the competition for attention from other sources of available stimuli. When the environment is lacking in stimulation individuals tend to pay more attention to bodily symptoms. Conversely, when an individual's attention is drawn to the external environment, bodily symptoms are less likely to be noticed. This finding has wide day-to-day applications ranging from why people cough in the boring parts of movies and lectures to explaining demographic differences in symptoms reports, such as increased symptom reporting among the socially isolated and the unemployed. It also has clinical applications in chronic pain and other chronic medical conditions where patients' isolation may exacerbate the condition by increasing preoccupation with symptoms.

Cognitive schemas can also strongly influence the reporting of physical symptoms by guiding the way individuals pay attention to their body. Schemas determine the organization of incoming information and guide health directed behaviour. There is a strong tendency for individuals to search for information that is consistent with existing schemas and disregard information that does not fit. Individuals also attach more importance to symptoms consistent with a current cognitive schema than other symptoms. Schemas may develop through personal experience with the condition or by having come across the illness through family, friends, or in the media. Illness schemas can vary from vague ideas about the types of symptoms that represent an illness to more elaborate and detailed conceptions of individual illnesses. Medical students' disease, where students studying a particular illness notice they also have the symptoms of the condition, and episodes of mass psychogenic illness are more dramatic demonstrations of this phenomenon, but the process is seen on a more subtle level with response to placebos (see below). Here, following treatment, a new cognitive schema may shift attention towards symptoms that indicate recovery rather than those of the illness.

Patient delay

There is growing research to suggest that patients' interpretation of their symptoms can influence help-seeking behaviour.[22] One medical condition where delay can have serious consequences is myocardial infarction, as early arrival at hospital is strongly associated with improved chances of survival. There is a large variation in how long patients delay before seeking help, and a strong predictor of early arrival at hospital is the belief that the symptoms are a heart attack.[23] Heart attacks are generally seen as sudden and dramatic events that involve severe chest pain and collapse. In the case of myocardial infarction patients, the mismatch between these expectations and the symptoms experienced gives rise to patient delay.

Research investigating the stages of patient delay for medical conditions has generally found three main stages prior to entering treatment, with each stage influenced by a different set of factors and decisional processes. The first interval is generally referred to as appraisal delay, which is the time period from when the individual first detects symptoms to when an illness is inferred. The main influences on this period are factors related to interpretation of symptoms. The second interval is called illness delay—the period from the time the individual decides he or she is ill until the decision is made to seek medical help. The final period called utilization delay is the time until the individual enters hospital or has contact with medical personnel. This first period of appraisal delay has been generally found to cause the largest contribution to overall delay.[24]

High health service users

A large percentage of medical consultations are made for non-medical complaints. This is particularly so for primary health care services. A number of studies have found that a small percentage of individuals without significant medical illness use a disproportionately large amount of medical services and at considerable cost.[25] These individuals have been variously labelled as somatizers, hypochondriacs, the worried well, patients with medically unexplained symptoms, and multiple attenders.

Research on high health service users suggests they are higher in trait anxiety. This is consistent with research showing a strong relationship between the somatic complaints and high levels of psychological distress or neuroticism. Individuals high in anxiety tend to be more introspective, watchful for any unusual symptoms, and develop more negative interpretations of symptoms they experience.[26] Symptoms of anxiety, such as tachycardia, can also be misinterpreted as signs of a physical illness by some patients.

Some individuals also seem to have a tendency to make catastrophic interpretations about physical symptoms and this may influence frequency of presentation to medical services and recovery from illness. Catastrophizing in pain patients has been associated with disability, and drop-out from pain-management programmes.[27] It has also been associated with higher levels of fatigue and disability in chronic fatigue syndrome patients.

Catastrophizing is also seen in 'cardiac invalidism'. Here patients adopt an extremely passive, dependent, and helpless role in the belief that any form of overly vigorous activity will bring on another myocardial infarction. A hypersensitivity to bodily symptoms means that normal sensations may be misconstrued to indicate overexertion or an impending fatal myocardial infarction. This pattern often results in a cycle of inactivity and loss of physical condition, which in turn can support these beliefs when patients exert themselves. Many patients who develop highly negative illness beliefs overuse medical services mainly for reassurance about symptoms.

The issue of reassurance in medical consultations is relevant here. One of the common patient expectations in primary care consultations is to have a better understanding of current symptoms. For many patients, being told there is no serious medical problem underlying their symptoms is effective in reducing concern about their condition, but for a significant number there remains worry about their health status. Continued anxiety in this group often results in further needless consultations and investigations. Evidence suggests that patients' existing beliefs about their condition are predictive of reassurance failure and that for reassurance to be effective, patients' concerns need to be elicited and appropriate information provided to explain either the patient's symptoms or why serious pathology has been ruled out.[28]

Recent work on improving reassurance following medical testing has suggested providing information to patients about the meaning of normal test results before testing, may weaken patients' preconceived beliefs about their condition and provide a context to help understand the test result. In this study, providing patients with information about normal test results prior to testing, improved their reassurance, reduced their symptoms, and lessened their use of unnecessary medication.[29]

If patients' ideas and beliefs about their symptoms are not addressed when symptoms persist or recur it is likely that health worry will also be reactivated as the patient still lacks a satisfactory cognitive model or explanation that enables them to interpret their symptoms as benign. A practical consequence of these findings is the need for clinicians to elicit the patient's own attributions and concerns about their symptoms and to use these as the basis for dealing with misconceptions and providing the patient with a more benign explanation of their symptoms.[29]

Cognitive models of illness

Research suggests patients cluster their ideas about an illness around five coherent themes or components, which health psychologists have called illness perceptions.[30] These provide a framework for patients to make sense of their symptoms, assess health risk, and direct action in the recovery phase. The major cognitive components are as follows.

- Identity: the label of the illness and the symptoms the patient views as being part of the disease.
- Cause: personal ideas about aetiology, which may include simple single causes or more complex multiple causal models.
- Time-line: the patient's belief about the likely time course of the illness (e.g. acute, chronic, or episodic).
- Consequences: expected impact of the illness on the patient's life.
- Cure/control: the patient's beliefs about the extent to which the illness is amenable to cure or control either through personal actions or by treatment.

These components show logical interrelationships. For example, a strong belief that the illness can be cured or controlled is typically associated with short perceived illness duration and relatively minor consequences.

The theoretical framework for this research is derived from the self-regulatory model developed by Leventhal *et al.*[30] This model views illness perceptions as critical in guiding the patient's coping efforts to deal with symptoms, illness, and threats to health. It consists of four components: the cognitive representation of the illness, the emotional response to the illness and treatment, the coping directed by the illness representation, and the individual's appraisal of the coping outcome.

Patient cognitive models of their illness are, by their nature, private. Patients' are often reluctant to discuss their beliefs about their illness in medical consultations because they fear being seen as ignorant or misinformed. Until recently, assessment of illness perceptions has been by open-ended interviews designed to encourage patients to elaborate their own ideas on the illness. However, questionnaires have been developed to measure illness perceptions in a variety of illnesses[31,32] as well as specific beliefs about medication.[33]

The illness perception approach has recently been applied to a large number of health conditions (see Hagger and Orbell[34] for a meta-analysis). Current research in this area is building on these findings to develop cognitive–behavioural interventions designed to modify dysfunctional illness perceptions and provide better recovery. A good example of this is a study showing that the early elicitation and modification of dysfunctional illness beliefs can improve recovery and return to function in patients with a recent myocardial infarction.[35]

Health care behaviour

In this section we examine the role of psychological processes in the delivery of health care by focusing on two broad areas: doctor–patient communication and health care in hospital.

Doctor–patient communication

There is now considerable evidence not only of patient dissatisfaction with medical communication but also of widespread non-compliance with subsequent treatment recommendations. Early research revealed that patient dissatisfaction was often associated with receiving insufficient information, poor understanding of the medical advice, and subsequent reluctance or inability to follow recommended treatment or advice. Another source of patients' dissatisfaction is the perception that the doctor lacks interest and empathy, and is unwilling to involve them in decision-making during the consultation. Thus, an overview of research in this area[36] revealed that patient satisfaction was higher following consultations in which the doctor engaged in more social conversation, positive verbal and non-verbal behaviour, and partnership building.

A range of frameworks have been developed for describing the process of the consultation. Similarly various methods have been devised for analysing the interactional processes which occur during the consultation[36] and Roter *et al.*[37] have used these

analyses to propose five distinct patterns of communication in doctors:

- narrowly biomedical, characterized by closed-ended medical questions and biomedical talk
- expanded biomedical, similar to the narrowly biomedical but with moderate levels of psychosocial discussion
- biopsychosocial, reflecting a balance of psychosocial and biomedical topics
- psychosocial, characterized by psychosocial exchange
- consumerist, characterized by patient questions and information giving by the doctor.

The highest levels of patient satisfaction were found with those who had seen doctors using the psychosocial communication pattern, whereas the lowest satisfaction scores were recorded in those who had experienced either of the two biomedical patterns.

An alternative and broader distinction has been made between consultations which are described as patient centred and those which are doctor centred, reflecting the extent to which the doctor or patient determines what is discussed. Doctor-centred consultations are ones in which closed questions are used more often and the direction is determined by the doctor, typically with a primary focus on medical problems. In contrast, patient-centred encounters involve more open-ended questions with greater scope for patients to raise their own concerns and agendas.

Patient satisfaction and understanding of their illness following the medical consultation can play a major role in influencing adherence with treatment or advice as well as other outcomes including health and well-being. A number of studies have demonstrated beneficial effects on patients' health and well-being arising from positive experiences in medical consultations.[38] These have focused on psychological states such as anxiety as well as changes in specific physical variables such as blood pressure and blood glucose control. Some of the most impressive findings here have been found in the patient-intervention studies, which are described below.

One important spin-off from the findings in this area has been the development of communication skills training packages for medical undergraduates and for experienced clinicians, particularly for improving skills in difficult areas of communication such as giving 'bad news'. There have also been a number of specific interventions aimed at patients. Generally, these have involved interventions for patients prior to a consultation in order to increase their level of participation, particularly to ensure that their own concerns are dealt with and that information provided by the doctor is clearly understood.

Health care in hospital

Patients experience many stressors in hospital and these arise from a range of factors including enforced lifestyle changes and the demands involved in developing good relations with hospital staff.[39] Other hospital stressors include worries about aspects of communication with staff, as well as concerns about investigations and treatment. Even such factors as the layout and colour of the ward, and the view from the patient's bed have been found to affect recovery. Not surprisingly, studies that have compared home-treated and hospitalized patients with the same condition have shown less psychosocial distress in those remaining at home.

In addition to these general psychological impacts of hospitalization, there may be specific problems or demands which occur either as a result of the particular health problem or the type of treatment which the patient has to undergo. An example of the way in which patients' health problems may influence their experience of hospital care can be seen in some of the studies of patients with HIV/AIDS who may experience negative or blaming attitudes from staff or other patients. For example, with AIDS patients being treated in either special care units or integrated in more general hospital settings, the latter group reported higher levels of stress associated with feelings of abandonment, and impersonal or discriminatory treatment. Where staff perceive patients as instrumental in having brought about their own condition through their own behaviour or neglect, they may be less committed, motivated, and sympathetic towards them.

A number of studies have been made of the psychological effects of specific treatment settings such as intensive care units (ICU) and haemodialysis units. Studies of patients in intensive care reveal high levels of psychological distress both during and for some time after their stay.[40] A range of factors seem to be involved, including being intubated and not being able to communicate. Even physical aspects of the ICU can have significant effects. Thus comparisons of patients in intensive care units with and without windows found that those in the windowless units were less well oriented during their stay and had a less accurate recall of their length of stay afterwards. In addition to these general problems associated with the intensive care units, other studies have assessed the degree of stress experienced by staff and visitors. For patients' relatives there is evidence that they find the time spent by the patient on life support in the intensive care unit particularly worrying. During this time they experience considerable fear and uncertainty but this can be improved by seeking information and the use of other resources.

In contrast with the acute psychological restrictions and demands of intensive care, some patients such as those on renal dialysis are subject to much more chronic restrictions. Dialysis can have major effects on an individual's psychological and social functioning, particularly giving rise to vocational impairment, reduced sexual activity, and mood changes.[41] In addition to the physical limitations and demands of dialysis, patients are also faced with the need to adhere to strict recommendations regarding diet and fluid consumption, as well as complex medication regimens. A number of aspects of dialysis can give rise to psychological distress, including the constant threat of death, dependence on the dialysis machine and medical staff. The stringent dietary and liquid restrictions are also important factors in patients' feelings of helplessness and lack of control. The ways in which patients cope can have important influences on their well-being and outcome. For example, problem-focused types of coping have been shown to be associated with better adherence to fluid intake restrictions, when these coping strategies were used in response to stressors arising from a relatively controllable aspect of dialysis. For those stressors, which patients perceived as less controllable, emotion-focused coping strategies provide better levels of adherence.

Many medical procedures in hospital can give rise to considerable discomfort and anxiety. These include certain treatments such as surgery, and specific investigative procedures such as barium radiography, endoscopy, and cardiac catheterization, which may not only be uncomfortable and sometimes physically distressing but which also carry the threat of uncovering a serious medical

condition.[42] Consequently a number of psychological interventions have been developed to prepare patients for surgery or other stressful procedures in the hospital setting. In broad terms they can help by providing the patient with information to reduce the uncertainty of the event, or with specific behavioural or cognitive skills to help with some of the discomfort or pain.[43]

These interventions have been found to improve a range of post-surgical outcomes, including anxiety, pain and use of pain medication, length of stay in hospital, and various indicators of recovery. All the interventions have been found to be successful in improving at least one aspect of outcome, and the majority of them have a positive impact on many of the outcomes. A meta-analysis by Johnston and Vogele[43] revealed that the largest recovery effects were obtained for pain, negative affect, and physiological indices of recovery but there was considerable variation in the magnitude of these effects. Smaller but more consistent advantages of psychological preparation were found on pain medication and length of hospital stay. The interventions, which had the most widespread overall effects on all the outcomes, were found to be procedural information provision and behavioural instructions. In addition to these specific psychological preparations, there is now evidence that the pre- and post-surgical social setting can have a significant effect on recovery. Studies have also revealed clear beneficial effects of sharing a room with someone who was recovering from surgery. Patients who had post-surgical room-mates, who had undergone the same type of surgery, have been shown to be less anxious prior to surgery, engaged in more post-surgical physical activity, and were discharged home sooner.

Treatment behaviour

Patients respond to their treatment in a range of ways and these can have very significant effects on clinical outcomes. Two major areas of patient behavioural variation are seen in the extent to which patients adhere to their prescribed treatment and in the non-specific or placebo effects of the treatment on clinical outcome. An overview of research on these two areas is now presented.

Adherence

The extent to which the patient adheres to the advice or treatment offered in health care consultations has been widely studied. Most medical consultations result in the prescription of treatment or advice, and the use of medicines is a key aspect to the self-management of most chronic illnesses. However, many patients fail to do this and low rates of adherence to recommended treatment are seen as problematic in chronic physical and psychiatric illnesses.[44]

The incidence of reported medication non-adherence varies greatly from 4 to 92 per cent across studies, converging at 30 to 50 per cent in chronic illness. In primary prevention studies, it has been found that many participants drop out of lifestyle change programmes, designed to improve diet or reduce health-risk behaviours. Even patients who have experienced major health problems, such as heart attacks, may show low levels of uptake of rehabilitation programmes as well as considerable variation in the adoption of recommended lifestyle change. In the area of mental health, there is also evidence of significant rates of non-adherence to various recommendations from health care providers.

Non-adherence behaviours may be categorized as either intentional or unintentional. Intentional non-adherence arises when the patient makes a strategic decision not to take the treatment as instructed. An example of this type of behaviour has been found among hypertensive patients who believed that they could judge when their blood pressure was high by the presence of symptoms such as stress or headache and thus took antihypertensive medication only when these symptoms were experienced. From a self-regulatory perspective, the level of treatment adherence may be indicative of a strategic coping response, which is entirely consistent with the patient's view of their problem. Thus, patients who believe that their problem will not last for long have been found to be less likely than those with a more chronic time-line representation to adhere to their medication over a long period of time.

Non-adherence may be unintentional when the patient's intentions to follow treatment recommendations are thwarted by barriers such as forgetting, and inability to follow treatment instructions because of a lack of understanding or physical problems such as poor eyesight or impaired manual dexterity. Thus, if the quality of communication is poor and patients receive information, which is difficult to understand or recall, as has been outlined above, then this makes it less likely that treatment will be adhered to.

The determinants of non-adherence

One very obvious explanation for non-adherence arises from poor understanding and recall of information presented in the medical consultation. Many patients lack basic knowledge about their medication but there is no simple relationship between this and their adherence. Reviews of adherence research fail to demonstrate a consistent positive association between knowledge and adherence. Moreover, interventions that enhance knowledge do not necessarily improve adherence. Patient satisfaction can act as a mediator between information provision, recall, and adherence since patient surveys reveal that many patients wanted more information than they were given. Dissatisfaction with attributes of the practitioner or the amount of information and explanation provided may act as a barrier to adherence by making the patient less motivated towards treatment.

The emphasis of adherence research over the last decade or so has moved away from attempts to identify stable trait factors which characterize the non-adherent patient to achieving a greater understanding of how and why patients decide to take some treatments and not others. Much of this research is informed by psychological theories, which conceptualize behaviour as the product of cognition which occurs within a social framework.

The application of the social cognition models, described earlier in this chapter, indicates that medication non-adherence may arise from a rational decision on the part of the patient and identifies some of the cognitions which are salient to these decisions. The types of beliefs and attitudes specified by such theories as the Health Belief Model, the Theory of Planned Behaviour and the Self-Regulatory Model (SRM) have all been used to explain aspects of treatment adherence. The SRM also acknowledges the importance of symptom perception in influencing illness representations and adherence as a coping behaviour. Confirmatory evidence for this is provided by findings from studies of patients with hypertension and with diabetes, both of whom commonly use perceived symptoms to indicate their blood pressure and glucose levels respectively, and to guide self-treatment. However, patients' beliefs about their symptoms and estimations of their own blood pressure

and glucose levels are often erroneous, and this can result in poor control of symptoms and illness.

More recent research has begun to focus on the role of people's beliefs about medicines and the ways in which these could influence adherence.[45] This research has revealed two broad factors describing people's beliefs about their prescribed medicines: their perceived necessity for maintaining health (specific-necessity) and concerns based on beliefs about the potential for dependence or harmful long-term effects and that medication taking is disruptive (specific-concerns). Two factors were also found to describe people's beliefs about medicines in general. The first relates to the intrinsic properties of medicines and the extent to which they are harmful addictive substances (general-harm) and the second factor comprises concerns that medicines are overused by doctors (general-overuse).

People's views about the specific medication regimen prescribed for them were found to be much more strongly related to adherence reports than are more general views about medicines as a whole. Moreover, interplay was found between concerns and necessity beliefs, which suggests that people engage in a risk–benefit analysis and consequently attempt to moderate the perceived potential for harm by taking less. Patients with stronger concerns based on beliefs about the potential for long-term effects and dependence reported lower adherence rates, whilst those with stronger beliefs in the necessity of their medication reported greater adherence to medication regimen.[45] This work points to the importance of accessing patients' beliefs as a prerequisite of any intervention designed to increase medication adherence. In particular, it would seem important to identify specific concerns about treatment and to allay these in ways which make sense to the patient.

The placebo response

The term 'placebo' is used to describe a treatment that gains a response due to its therapeutic intent rather than the specific ingredients of the treatment itself. Placebo responses have been shown for a wide variety of medical treatments including surgery, psychotherapy, medication, therapeutic ultrasound, injections, and aerosol sprays. Placebos have also been demonstrated to have effects in countless medical conditions and also on a number of physiological functions such as blood pressure, heart rate, gastric motility, lung function, and postoperative swelling. Adverse effects from placebos or so-called 'nocebo' effects have also been noted in the literature.

Characteristics of the treatment itself and the setting it is administered in can have a strong influence on the magnitude of the placebo response. In general, treatments that involve more serious rituals and sophisticated equipment such as surgery have stronger placebo effects. Likewise, other treatments imparting a powerful impression to the patient such as foul- or strong-tasting medicine, injections, and precise instructions also enhance the placebo response. The colour of medication has been shown to have some effect depending on the condition, and known brand names seem to have an edge in placebo response over unknown drug companies.

A similar theme runs through the clinician characteristics that increase the placebo response. Clinicians and clinics seen as having high status and having high levels of credibility have an improved placebo response. At the same time, the doctor–patient interaction is also important. If the doctor shows high levels of concern and empathy for the patient then the response increased. High confidence shown by the doctor in the treatment administered to the patient along with a clear indication of the expected response of the treatment is also likely to improve the likelihood of placebo response.

In contrast, isolating characteristics of patients who are placebo responders has yielded inconclusive results. Much of the evidence for the role of demographic, intellectual, or personality characteristics of patients likely to respond to placebos is mixed and inconsistent. Studies have found individuals who responded to placebos in one setting to be unresponsive in another. Likewise, conditioning studies have shown individuals who have been unresponsive can later respond. These findings point to the fact that individual characteristics probably play a less significant role than situational factors and the doctor–patient interaction in influencing the placebo response.

Treatment response has been divided into specific and non-specific components, with the non-specific component encompassing factors such as clinician attention, expectation, reputation, treatment setting, etc. Determining the magnitude of the specific and non-specific components of medical treatment is a difficult and probably impossible task. In an attempt to determine how powerful non-specific effects are under ideal circumstances, Roberts *et al.*[46] chose to look at the effect of a number of medical treatments later shown to be ineffective but where clinician faith in the treatment was initially positive. Pooling the data from diverse treatments such as gastric freezing for duodenal ulcer and glomectomy for asthma, this study found 40 per cent of patients had an excellent response to the treatments, 30 per cent good, and 30 per cent poor. This suggests that under ideal circumstances where clinician and patient expectations are high and the treatment is administered in a credible way, non-specific factors can by themselves exert a powerful effect.

The role of compliance with placebos also appears to be important. In a review of five placebo-controlled studies measuring both compliance with medication and outcome, Epstein[47] found that subjects who were more compliant did better on outcome measures, regardless of whether they were on placebo or active treatment. Outcomes included prevention of relapse in schizophrenia, reduction of fever or infection in cancer patients, alcohol abstinence, reduction of weight, and prevention of mortality in patients with heart disease. In a later study, Horwitz *et al.*[48] also found the risk of death was substantially less in patients who took more than 75 per cent of their medication regardless of whether the medication was placebo or β-blocker. This suggests the act of compliance may have some effects of other health-promoting behaviours or cause cognitive or emotional changes that may influence health in the long-term.

There have been a number of theories proposed to explain the placebo response but no one theory yet provides an adequate integrated theoretical framework. The reduction of anxiety following treatment and consequent effect on symptoms has been proposed as one mechanism by which the placebo effect may operate, but changes in anxiety states have not reliably been associated with placebo responses. The role of the medical situation and its accoutrements being associated through classical conditioning with symptom relief is likely to play some role. There is, however, little direct research on this proposed mechanism, although

classical conditioning of drug responses has been shown in certain situations.

Two other theories proposed have been the role of cognitive dissonance and patient expectations. The cognitive dissonance argument proposes that the placebo effect may be due to the pressure on individuals to show consistency in their views and actions. Therefore, for some individuals, having treatment is inconsistent with not showing any change in symptoms and this may encourage the person to reduce this inconsistency. The role of patient expectations and placebo effects is an area that has not received a great deal of systematic research. It is suggested that patient expectations may cause changes in cognitive schemas that influence the types and nature of symptoms that patients pay attention to following treatment. New developments in research on illness perception and beliefs about treatment, outlined above, as well as in the fields of neurobiology and neuroimaging[49] hold considerable promise for increasing our understanding of the nature of the placebo effect and its determinants.

Conclusions

This selective overview of health psychology has demonstrated the range of psychological processes in health, illness, and health care. At the present time it is primarily a disciplinary area of psychology with an emphasis on research into health and illness behaviour. However, many interventions have been developed for healthy individuals, patients, and health care staff. This practitioner aspect of health psychology is now being accompanied by specific professional developments, and formal postgraduate training in health psychology is now available in many countries.[50]

Health psychology has established itself rapidly but it is still very much an emerging discipline. Greater insights are needed into the ways in which psychological processes can influence health and illness, and more comprehensive models are required for explaining all aspects of health and illness behaviour. In the long term this will result in the increasing use of psychological interventions for preventing and managing health problems and for the effective delivery of health care.

Further information

Weinman, J., Johnston, M., and Molloy, G. (eds.) (2007). *Health psychology, 4 volume set*. Sage, London.

http://www.health-psych.org/

Ayers, S., Baum, A. McManus, I.C. (eds.), *et al.* (2007). *Cambridge handbook of psychology, health and medicine* (2nd edn). Cambridge University Press, Cambridge.

Johnston, M., Weinman, J., and Wright, S. (1995). *Health psychology: an assessment portfolio*. NFER Nelson, Windsor.

References

1. Mokdad, A.H., Marks, J.S., Stroup, D.F., *et al.* (2004). Actual causes of death in the United States, 2000. *JAMA*, **291**, 1238–45.
2. Conner, M. and Norman, P. (eds.) (2005). *Predicting health behaviour*. Open University Press, Buckingham.
3. Petrie, K.J. and Weinman, J. (eds.) (1997). *Perceptions of health and illness: current research and applications*. Harwood Academic, London.
4. Ayers, S., Baum, A., Newman, S. (eds.), *et al.* (2007). *Cambridge handbook of psychology, health and medicine* (2nd edn). Cambridge University Press, Cambridge.
5. Ogden, J. (2007). *Health psychology: a textbook* (4th edn). Open University Press, Buckingham.
6. McEwen, B.S. (1998). Protective and damaging effects of stress mediators. *New England Journal of Medicine*, **338**(3), 171–9.
7. Segerstrom, S.C. and Miller, G.E. (2004). Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological Bulletin*, **130**(4), 601–30.
8. Cohen, S., Tyrell, D.A., and Smith, A.P. (1993). Psychological stress and susceptibility to the common cold. *New England Journal of Medicine*, 325, 606–12.
9. Kiecolt-Glaser, J.K., Marucha, P.T., Malarkey, W.B., *et al.* (1995). Slowing of wound healing by psychological stress. *Lancet*, **346**, 1194–6.
10. Broadbent, E., Petrie, K.J., Booth, R., *et al.* (2003). Psychological stress impairs early wound repair following surgery. *Psychosomatic Medicine*, **65**, 865–9.
11. Friedman, H.S. and Booth-Kewley, S. (1987). The "disease-prone personality": a meta-analytic view of the construct. *The American Psychologist*, **42**(6), 539–55.
12. Hecker, M.H.L., Chesney, M.A., Black, G.W., *et al.* (1988). Coronary-prone behaviors in the western collaborative group study. *Psychosomatic Medicine*, **50**, 153–64.
13. Thoresen, C., Friedman, M., Powell, L.H., *et al.* (1985). Altering the Type A behavior pattern in postinfarction patients. *Journal of Cardiopulmonary Rehabilitation*, **5**, 258–66.
14. Denollet, J., Pedersen, S.S., Vrints, C.J., *et al.* (2006). Usefulness of type D personality in predicting five-year cardiac events above and beyond concurrent symptoms of stress in patients with coronary heart disease. *The American Journal of Cardiology*, **97**(7), 970–3.
15. Wiebe, D. and Smith, T.W. (1997). Personality and health. In *Handbook of personality psychology* (eds. R. Hogan and J.A. Johnson), pp. 891–918. Academic Press, San Diego, CA.
16. Carver, C.S., Smith, R.G., Antoni, M.H., *et al.* (2005). Optimistic personality and psychosocial well-being during treatment predict psychosocial well-being among long-term survivors of breast cancer. *Health Psychology*, **24**(5), 508–16.
17. Adler, N.E., Boyce, T., Chesney, M.A., *et al.* (1994). Socioeconomic status and health: the challenge of the gradient. *The American Psychologist*, **49**(1), 15–24.
18. Schoenborn, C.A. (1993). The Alameda Study—25 years later. In *International review of health psychology*, Vol.**2** (eds. S. Maes, H. Leventhal, and M. Johnston), pp. 81–116. John Wiley & Sons Ltd., Chichester.
19. Warburton, D.E.R., Nicol, C., and Bredin, S.D.S. (2006). Health benefits of exercise: the evidence. *CMAJ: Canadian Medical Association Journal*, **174**, 801–9.
20. West, R.J. (2005). Time for a change: putting the transtheoretical (stages of change) model to rest. *Addiction*, **100**(8), 1036–9.
21. Pennebaker, J.W. (1982). *The psychology of physical symptoms*. Springer-Verlag, New York.
22. Cameron, L., Leventhal, E.A., and Leventhal, H. (1995). Seeking medical care in response to symptoms and life stress. *Psychosomatic Medicine*, **57**, 37–47.
23. Horne, R., James, D., Petrie, K.J., *et al.* (2000). Patients' interpretation of symptoms as a cause of delay in reaching hospital during acute myocardial infarction. *Heart*, **83**, 388–93.
24. Andersen, B.L. and Cacioppo, J.T. (1995). Delay in seeking a cancer diagnosis: delay stages and psychophysiological comparison processes. Special issue: social psychology and health. *The British Journal of Social Psychology*, **34**, 33–52.
25. Verhaak, P.F.M., Meijer, S.A., Visser, A.P., *et al.* (2006). Persistent presentation of medically unexplained symptoms in general practice. *Family Practice*, **23**, 414–20.
26. Sensky, T., Macleod, A.K., and Rigby, A.F. (1996). Causal attributions about common somatic sensations among frequent general practice attenders. *Psychological Medicine*, **26**, 641–6.

27. Sullivan, M.J.L., Thorn, B., Haythornthwaite, J.A., *et al.* (2001). Theoretical perspectives on the relation between catastrophising and pain. *The Clinical Journal of Pain*, **17**, 52–64.
28. Donkin L, Ellis, C.J., Powell, R., *et al.* (2006). Illness perceptions predict reassurance following negative exercise testing result. *Psychology and Health*, **21**, 421–30.
29. Petrie, K.J., Müller, J.T., Schirmbeck, F., *et al.* (2007). Effect of providing information about normal test results on patients' reassurance: randomized controlled trial. *British Medical Journal*, **334**, 352–5.
30. Leventhal, H., Nerenz, D.R., and Steele, D.J. (1984). Illness representations and coping with health threats. In *A handbook of psychology and health*, Vol. 4 (eds. A. Baum and J. Singer), pp. 219–52. Erlbaum, Hillsdale, NJ.
31. Weinman, J., Petrie, K.J., Moss-Morris, R., *et al.* (1996). The illness perception questionnaire: a new method for assessing illness perceptions. *Psychology and Health*, **11**, 431–46.
32. Broadbent, E., Petrie, K.J., Main, J., *et al.* (2006). The brief illness perception questionnaire (BIPQ). *Journal of Psychosomatic Research*, **60**, 631–7.
33. Horne, R., Weinman, J., and Hankins, M. (1999). The beliefs about medicines questionnaire: a new method for assessing cognitive representations of medication. *Psychology and Health*, **14**, 1–24.
34. Hagger, M.S. and Orbell, S. (2003). A meta-analytic review of the common-sense model of illness representations. *Psychology and Health*, **18**, 141–84.
35. Petrie, K.J., Cameron, L.D., Ellis, C.J., *et al.* (2002). Changing illness perceptions after myocardial infarction: an early intervention randomized controlled trial. *Psychosomatic Medicine*, **64**, 580–6.
36. Roter, D. and Hall, J.A. (1989). Studies of doctor patient interaction. *Annual Review of Public Health*, **10**, 163–80.
37. Roter, D., Stewart, M., Putnam, S.M., *et al.* (1997). The patient-physician relationship: communication patterns of primary care physicians. *The Journal of the American Medical Association*, **277**, 350–6.
38. Stewart, M.A. (1995). Effective physician–patient communication and health outcomes: a review. *Canadian Medical Association Journal*, **152**, 1423–33.
39. Koenig, H.G., George, L.K., Stangl, D., *et al.* (1995). Hospital stressors experienced by elderly medical in-patients: developing a hospital stress index. *International Journal of Psychiatry in Medicine*, **25**, 103–22.
40. Andrews, P., Azoulay, E., Antonelli, M., *et al.* (2005). Year in review in intensive care medicine, 2004. III. Outcome, ICU organization, scoring, quality of life, ethics, psychological problems and communication in the ICU, immunity and hemodynamics during sepsis, pediatric and neonatal critical care, experimental studies. *Intensive Care Medicine*, **31**, 356–72.
41. Kimmel, P. (2000). Psychosocial factors in end-stage renal disease patients treated with hemodialysis: correlates and outcomes. *American Journal of Kidney Diseases*, **35**, S132–40.
42. Weinman, J. and Johnston, M. (1988). Stressful medical procedures: an analysis of the effects of psychological interventions and of the stressfulness of the procedure. In *Topics in health psychology* (eds. S. Maes, P. Defares, I.G. Sarason, and C.D. Speilberger), pp. 205–17. Wiley, Chichester.
43. Johnston, M. and Vogele, C. (1993). Benefits of psychological preparation for surgery: a meta-analysis. *Annals of Behavioral Medicine*, **15**, 245–56.
44. Osterberg, L. and Blaschke, T. (2005). Adherence to medication. *The New England Journal of Medicine*, **353**, 487–97.
45. Horne, R. and Weinman, J. (1999). Patients' beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. *Journal of Psychosomatic Research*, **47**, 555–67.
46. Roberts, A.H., Kewman, D.G., Mercier, L., *et al.* (1993). The power of non-specific effects in healing: implications of psychosocial and biological treatments. *Clinical Psychology Review*, **13**, 375–91.
47. Epstein, L.H. (1984). The direct effects of compliance on health outcome. *Health Psychology*, **3**, 385–93.
48. Horwitz, R.I., Viscoli, C.M., Berkman, L., *et al.* (1990). Treatment adherence and risk of death after a myocardial infarction. *Lancet*, **336**, 542–5.
49. Benedetti, F., Mayberg, H.S., Wager, T.D., *et al.* (2005). Neurobiological mechanisms of the placebo effect. *The Journal of Neuroscience*, **24**(45), 10390–402.
50. Jansen, M. and Weinman, J. (eds.) (1991). *The international development of health psychology*. Harwood Academic, London.

5.7

The organization of psychiatric services for general hospital departments

Frits J. Huyse, Roger G. Kathol, Wolfgang Söllner, and Lawson Wulsin

Introduction

The organization of psychiatric services for general hospital departments might change in far-reaching ways in the coming decades. Whereas the focus was primarily on reactive services for inpatients on medical and surgical wards, the future should focus on more proactive integrated service delivery for the complex medically ill. The essential difference from other psychiatric services is that the population served is taken care of by medical specialists in the general health setting. Consequently services are delivered in the context of the medical-psychiatric interface. Consult requests are always formulated in this perspective: the patient is treated for a medical illness or physical complaints and there are signs of an interfering psychiatric disorder.[1] Nowadays these patients are referred to as the 'complex medically ill'.[1] Therefore triage and treatment integrated in the medical context is the area of expertise of consultation-liaison (CL) psychiatrists.

The development of this area of psychiatry has been hampered by dysfunctional splits in health care, such as between general and mental health care, both on the level of its organization as well as its reimbursement.[2, 3] Recent reports, such as the report of the joint working group of the United Kingdom Royal College of Physicians and the Royal College of Psychiatrists, which describe the psychological needs of the medically and surgically ill, provide guidance to counteract these dysfunctional splits.[4] As the delivery of care-trajectories for comorbid patients becomes more and more an issue on the health care agenda, CL psychiatrists should seize this opportunity and become advocates for integrated service delivery for the complex medically ill.

[1] Whereas in mental health comorbidity refers to making more than one criteria based psychiatric diagnosis, in the CL literature the term 'comorbidity' is generally used to describe the combination of physical diseases and psychiatric disorders.

Current levels of service delivery

Around 1990 the extent of inpatient CL psychiatric service delivery was evaluated, based on the records of a representative national sample of hospitals (United States)[5] and based on a prospective multicentred study (Europe).[6] Both studies reported an average consult rate of 1 per cent, ranging up to 4–5 per cent in some university settings. This rate is much lower than the prevalence of psychiatric disorders in medical populations.[6] Taking this underutilization into account, the most striking finding was still the large variation in departments served and types of patients seen. The European Consultation-Liaison Workgroup's (ECLW) Collaborative Study made clear that CL psychiatric service delivery is primarily an emergency service. Most referrals were late, as reflected by an average time of 11 days after admission before patients were referred. In addition one-third was emergency referrals: 'See the patient the same day'.[6, 7] Exceptions were the German psychosomatic services driven by their primary interest in patients with unexplained physical complaints and problems of coping with somatic illness using a more integrated liaison approach. These services showed higher consultation rates (between 2 and 4 per cent), provided more follow-up visits, and communication with aftercare providers.[6, 7–9]

It is now evident that mental disorders and physical diseases cluster in vulnerable patients. The prevalence of mental disorders in the general hospital population is on average twice as high compared to that of the general population. However, when focusing on specific populations such as cardiac, diabetes, or transplantation, the rates of major depression may reach up to 30 per cent[10, 11] (see other chapters of this section). Patients in the general hospital setting are primarily treated for their physical diseases. However, the multiple interactions between the comorbid medical and psychiatric disorders make them complex. This justifies an integrated approach and requires individualized multimodal and multidisciplinary care.[12,13] These complex patients are the target population for CL psychiatrists. They are in need of integrated services.

Table 5.7.1 Types of service delivery

1 Emergency services
• Attempted suicide
• Acute behavioural disturbances and their prevention
  • Deliria
  • Withdrawal

2 Regular consults for patients with possible interfering psychiatric complications, such as anorexia, factitious disorder, anxiety- or depressive disorders, adjustment disorders, somatization and organic mental disorders.
3 Integrated services • Participation in multidisciplinary clinics, such as pain, memory, or transplant • Participation in multi-disciplinary rounds on 'liaison-wards' or of disease management programmes, such as for patients with Parkinson disease, diabetes, cancer, or chronic heart failure • Screening for depression or complexity in at risk populations, including the development of related care trajectories • Clinical services for highly complex patients with both medical and psychiatric acuity, such as the medical psychiatric unit

Types of service delivery

Here several models of service delivery are described (Table 5.7.1). The models have an increasing level of sophistication determined by their level of integration and the related procedural collaborative activities. Service delivery requires by definition, negotiations with health plans for their reimbursement. This is especially true for the integrated models of service delivery.[7,13,14]

Consultations

Consultations are the classical mechanism for doctors to involve other medical specialists in the treatment of patients with additional medical problems. Patients are referred if the treating physician recognizes psychiatric comorbidity or a psychological problem and if he or she thinks that psychiatric evaluation and/or intervention may be helpful. The problem linked with this type of service delivery is that physicians often do not recognize psychiatric disturbance in medical patients.[15] In some cases, this problem is avoided by organizing a 'contract type' of consultation where every patient with a defined clinical problem is referred, for instance patients with attempted suicide.

Liaison[2]

Whereas in the consultation function psychiatrists wait for the referral, the liaison function is proactive. A preventive approach is implemented through weekly multidisciplinary rounds. In orderto establish such a role the consultant and a departmental head formulate a liaison arrangement for the provision of psychiatric services for a certain population, clinic or ward. An important additional aspect of the liaison model is its educational focus. Though every consult offers an educational opportunity, in the liaison function the consultant is better equipped to enhance the skills of the teams through weekly attendance of clinical rounds. Currently, the liaison model is restricted to tertiary care hospitals with more extensive CL psychiatric services. In the European collaborative study only 5 per cent of the consults came from a liaison arrangement.[6,8]

Psychiatric-medical, medical-psychiatric units, or psychosomatic units

The 'Psych-Med unit' is an integrated clinical service for high complex patients with unstable medical disease, such as diabetes, and psychiatric disorders. Due to their mutual interactions such patients require not sequential but integrated assessment and treatment, including both intensive medical and psychiatric nursing. Depending on the required acuity levels of physical and psychiatric nursing, different types of Psych-Med units can be described.[14] Dual-trained or combined staffs are selected to provide these levels of integrated care. Organizational prototypes of this function are the US-initiated psychiatric-medical clinics and the Germfan psychosomatic wards, which focus on adjustment disorders in medical patients and complex somatization.[14] During the last years, the efforts to improve integration between inpatient and outpatient care for complex patients led to the implementation of multidisciplinary, and specialized integrated treatment programmes for specific patient groups in day hospitals (e.g. for chronic pain patients and geriatric patients).

Screening

As the selection criteria for patients in the liaison function are not operationalized, referrals are intuitively generated on the basis of clinical expertise.[15] Nowadays instruments are available to support both clinical work and research. The liaison function can be seen as a precursor of screening. It will gradually merge into more structured preventive functions, defined by the needs of a target population and guided by screening.[16] Currently, two lines of screening are in its development. First of all there is a model with a primary focus on psychopathology and primarily depression, using the patient health questionnaire (PHQ) as an indicator of psychiatric comorbidity. It is used in (elderly) patients with physical disorders, such diabetes, or physical complaints of unknown origin.[11,17] Until now this is mainly used in an outpatient setting. The other approach taken is screening for 'complexity'. A European group has taken the approach to operationalize complexity and to develop a screener and an assessment tool to detect and analyse the complex medically ill.[12,16] The Complexity Prediction Instrument (COMPRI)—the screener—is to be applied at admission on an internal medicine ward to detect patients at risk for negative outcomes of care. At the same time a comparable instrument has been developed for the elderly population to detect patients who are frail and are or have an increased risk of becoming complex.[18] Other indicators of complexity such as administrative and clinical are discussed elsewhere.[12] The INTERMED-method has been developed for complexity assessment and the design of related integrated interventions.[19] It starts with a structured interview evaluating 16 health risk variables (Table 5.7.2). The fitting of these 16 risks with 4 prognostic variables in a biopsychosocial schema and the uniform-scoring system providing different levels of action visualized in different colours, supports decision-making and facilitates interdisciplinary communication. The integrated multidisciplinary interventions designed might require case-management.

[2] Here the term 'liaison' is only used to describe the specific 'liaison function' in addition to the basic consultation function.

Table 5.7.2 Health risks evaluated for complexity assessment with INTERMED-method

Chronicity	Is patient known with physical illness/disease
Diagnostic dilemma	Were physical symptoms clarified
Severity of symptoms	Physical functioning
Diagnostic challenge	Complexity of current medical problem
Restrictions in coping	Interferences of coping with medical problems
Psychiatric dysfunctioning	Psychiatric history
Resistance to treatment	Capacity to collaborate with treatment
Psychiatric symptoms	Severity of symptoms
Restrictions in integration	Social integration reflected by work and leisure
Social dysfunctioning	Quality of relations
Residential instability	Stability of housing
Restrictions of network	Availability of help
Intensity of treatment	Utilization
Treatment experience	Trust in health professionals
Organization of care	Participating health professionals
Appropriateness of referral	Capacity to deliver appropriate care

Until now this is an area of health care in which mental health care is not formally integrated. It is to be expected that screening for psychopathology or complexity in the chronically ill will be become important future tools to initiate integrated care and allow CL psychiatric teams to actively contribute to the care of complex patients.

The organization of a consultation-liaison psychiatric service

It is unrealistic to assume that in the future the needs of general hospital patients with psychiatric morbidity can be met simply by increasing staff. To see all patients with psychiatric comorbidity would require many times the present staffing levels. Consequently, CL psychiatrists should plan their services carefully together with their medical colleagues. The following points should be considered.

The population to be seen

Every consultant working in a general hospital, in primary care, or in a nursing-home setting needs to define what patients have to be seen and what services are to be delivered. In a general hospital, emergency services will be required for patients who attend for psychological reasons, including attempted suicide, and for patients with substance abuse withdrawal and acute deliria (Table 5.7.1). In addition consults should be done for patients with unexplained physical problems and other complex illness behaviour. The consult capacity beyond these two consult categories should be used for the development of more preventive consults integrated in existing forms of multidisciplinary service delivery. Selection of areas of interest depends on several factors, such as service delivery or research priorities of the hospital (for instance transplant or oncology) or an own research agenda. In primary care the target population in addition to the chronically ill will be patients with somatization problems including affective and anxiety disorders (see Chapter 5.2.3).

Psychiatric assessment

As in other settings, the formal psychiatric assessment is a crucial part of the services delivered. Specific to the setting is the differential diagnosis with physical disorders, the role and the meaning of physical deregulations, effects of pharmaca, the effects of the psychiatric disturbance on compliance with the treatment of the physical disease, the consequences of the assessment for the integrated prognosis, and the subsequent long-term integrated management of the patient. As the outcome of psychiatric disorders is clearly related to the interfering problems, which contribute to the complexity of patients, inclusion of the assessment of potential risks for such problems should be considered.[19]

Disciplines and staffing

The size and composition of the CL team needs to be defined depending on the size and type of the hospital and of the target group of patients (Table 5.7.1) as well on the financial possibilities and other available services. For the basic function, the assessment and treatment of patients seen for attempted suicide, one is referred to Chapter 4.15.4. In a European study (ECLW Collaborative Study) it became clear that there was a variation in team composition from monodisciplinary (medical model) to multidisciplinary (mental health model) depending on the size of the service as well as the country.[8] In addition to individual psychotherapeutic treatment, mutual adjustment with and instructions of other caretakers is a key aspect of CL work. Consequently, nowadays CL psychiatry cannot provide optimal care without team members focusing on psychological treatment and the organization of case-management required for long-term individualized care-trajectories. Good evidence is becoming available that psychological interventions (cognitive–behavioural, problem solving, and interpersonal psychotherapies) are effective in patients with physical illness and depression as well as unexplained physical complaints;[20,21] (see also Chapter 5.2.3). The effectiveness of interventions of CL psychiatric nurses depends on their roles. For the effectiveness of case-management in these patients is less evidence according to a recent systematic review.[22] Turning it another way around, as CL psychiatric nurses will often work in the chronic medically ill, a recent review has provided an overview to assess the effective elements of chronic disease management (Table 5.7.3).[23] To be able to contribute to integrated care programmes, such as for diabetes or for haemodialysis, tertiary care hospitals should have teams with, on average, one full-time equivalent of psychiatric staff per 300 beds and a secretary, in addition to psychiatric residents, nurses, and psychologists.[24] Both psychiatric and complexity screening functions need to be supported by manpower to translate the findings into clinical action, such as the design and implementation of a long-term individualized care trajectory and prevent decompensation in those who are vulnerable.

Relationship between medical staff, hospital board, and regional mental health facilities

For the development of more integrated services beyond emergency consultations good working relationships are required. The psychiatrist should be a formal member of the medical staff. Negotiations on the size and focus of CL psychiatric service delivery

Table 5.7.3 Evidence for effective chronic care management

There is evidence to support the following initiatives
- Broad chronic care management models
- Integrated community and hospital care
- Greater reliance on primary care
- Identifying people at greatest risk of complications and hospitalizations
- Involving people with long-term conditions in decision-making
- Providing accessible structured information for people with long-term conditions and their families
- Self-management education
- Self-monitoring and referral systems
- Electronic monitoring and telemonitoring
- Using nurse-led strategies, where appropriate

There is less evidence to support the following initiatives
- Case-management
- Evidence-based care pathways
- Shared learning among health professionals

There is limited information about
- New models of commissioning services
- Appropriate data collection and monitoring
- Linking health services with voluntary and community services

(From D. Singh (2006), *Transforming chronic care*, NHS Surrey and Suffolk UK. © Crown copyright.)

should be organized with representatives of the general hospital board, the regional mental health provider, and the health plan. They should decide on functions, budgets, and facilities. The lack of medical facilities in mental health institutions is a good reason to include the need to develop a psychiatric medical unit in the general hospital to serve the more serious medically ill from mental health institutions. Wards with both a psychiatric and medical function can solve the problems created by the artificial division between general and mental health care.

Audit

Both for financial purposes as well as for strategic planning practice audit is required. It does not make much sense to have an extensive audit system such as used in studies, unless this is used for projects. Otherwise an audit system integrated in the hospital mainframe seems the most appropriate, including the basic patient documentation, the reason for referral, the referring department, their diagnoses, and treatment.[25]

Training

In the 'western' world CL psychiatry is becoming more and more an area of special interest, which is reflected in a subspecialty-status in several countries, such as the United States, the United Kingdom, and Australia. Training should include specific medicopsychiatric aspects of the work, including psychopharmacology in the medically ill. Guidelines have been formulated by several associations and have been published.[26, 27]

Further information

Associations for consultation-liaison psychiatry exist, which organizational format differs by country. Since the first decade of the twenty-first century there is an increasing international exchange between leading organizations as well as among leaders in the field. Leading associations are

- The Academy of Psychosomatic Medicine in the USA: www.apm.org. This organization has an international membership and focus and
- The European Association of Consultation and Liaison Psychiatry and Psychosomatics: www.eaclpp.org

References

1. Huyse, F.J. and Stiefel, F.C. (eds.) (July 2006). *Integrated care for the complex medically ill*. Medical Clinics of North America, Elsevier.
2. Institute of Medicine. (2001). *Crossing the quality chasm: a new health system for the 21st century. Committee on quality of health care in America*. National Academy Press, Washington, DC.
3. Institute of Medicine. (2005). *Improving the quality of health care for mental and substance-use conditions: quality chasm series*. National Academy Press, Washington, DC.
4. Royal College of Physicians and Royal College of Psychiatrists. (1995). *The psychological care of medical patients. Recognition of need and service provision. Council report CR35*. Royal College of Physicians and Royal College of Psychiatrists, London.
5. Wallen, J., Pincus, H.A., Goldman, H.H., *et al.* (1987). Psychiatric consultations in short-term general hospitals. *Archives of General Psychiatry*, **44**, 163–8.
6. Huyse, F.J., Herzog, T., Lobo, A., *et al.* (2001). Consultation-liaison psychiatric service delivery: results from a European study. *General Hospital Psychiatry*, **23**(3), 124–32.
7. Huyse, F.J., Herzog, T., Lobo, A., *et al.* (2000). European consultation-liaison psychiatric services: the ECLW collaborative study. *Acta Psychiatrica Scandinavica*, **101**, 360—6.
8. Herzog, T., Creed, F., and Huyse, F.J., *et al.* (1994). Psychosomatic medicine in the general hospital. In *Psychiatry in Europe: directions and developments* (eds. C. Katona, S. Montgomery, and T. Sensky), pp. 143–51. Gaskell, London.
9. de Cruppe, W., Hennch, C., and Buchholz, C., *et al.* (2005). Communication between psychosomatic C-L consultants and general practitioners in a German health care system. *General Hospital Psychiatry*, **27**(1), 63–72.
10. Kathol, R., Saravay, S.M., Lobo, A., *et al.* (2006). Epidemiologic trends and costs of fragmentation. *The Medical Clinics of North America*, **90**, 549–72.
11. Egede, L.E. (2006). Disease-focussed or integrated treatment: diabetes and depression. *The Medical Clinics of North America*, **90**, 627–46.
12. De Jonge, P., Huyse, F.J., and Stiefel, F.C. (2006). Case and care complexity in the medically ill. *The Medical Clinics of North America*, **90**, 679–92.
13. Smith, G.C. and Clarke, D. (2006). Assessing the effectiveness of integrated interventions: terminology and approach. *The Medical Clinics of North America*, **90**, 533–48.
14. Wulsin, L.R., Söllner, W., and Pincus, H.A. (2006). Models of integrated care. *The Medical Clinics of North America*, **90**, 647–77.
15. Söllner, W., DeVries, A., and Steixner, E., *et al.* (2001). How successful are oncologists in identifying patient distress, perceived social support, and need for psychosocial counselling? *British Journal of Cancer*, **84**, 179–85.
16. Huyse, F.J., Stiefel, F.C., and de Jonge, P. (2006). Identifiers, or "red flags" of complexity and need for integrated care. *The Medical Clinics of North America*, **90**, 703–12.
17. Kroenke, K. and Rosmalen, J.G.M. (2006). Symptoms, syndromes, and the value of psychiatric diagnostics in patients who have functional somatic disorders. *The Medical Clinics of North America*, **90**, 603–26.
18. Slaets, J.P.J. (2006). Vulnerability in the elderly: frailty. *The Medical Clinics of North America*, **90**, 593–601.

19. Stiefel, F.C., Huyse, F.J., Söllner, W., *et al.* (2006). Operationalizing integrated care on a clinical level: the INTERMED project. *The Medical Clinics of North America*, **90**, 713–58.
20. Katon, W.J., Von Korff, M., Lin, E.H., *et al.* (2004). The pathways study: a randomized trial of collaborative care in patients with diabetes and depression. *Archives of General Psychiatry*, **61**(10), 1042–9.
21. Unutzer, J., Katon, W., and Callahan, C.M., (2002). Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *The Journal of the American Medical Association*, **288**(22), 2836–45.
22. Latour, C.H., van der Windt, D.A., de Jonge, P., *et al.* (2007). Nurse-led case management for ambulatory complex patients in general health care: a systematic review. *Journal of Psychosomatic Research*, **62**(3), 385–95.
23. Singh, D. (2006). *Transforming chronic care*. NHS Surrey and Suffolk, UK.
24. Herzog, T., Stein, B., Söllner, W., *et al.* (2002). *Practice guidelines for consultation-liaison psychosomatics*. Schattauer, Stuttgart.
25. Söllner, W., Stein, B., and Hendrischke, A., *et al.* (2005). A documentation form for the consultation-liaison services: development of the CL-BaDo. *Zeitschrift fur Psychosomatische Medizin und Psychotherapie*, **51**, 310–22.
26. Gitlin, D.F., Schindler, B.A., Stern, T.A., *et al.* (1996). Recommended guidelines for CL psychiatric training in psychiatry residency programs: a report from the academy of psychosomatic medicine task force on psychiatric training in CL psychiatry. *Psychosomatics*, **37**, 3–11.
27. Söllner, W., Creed, F. and the EACLPP Workgroup on training in CL. (2007). European guidelines for training in CL psychiatry and psychosomatics. *Journal of Psychosomatic Research*, **62**(4), 501–09.

SECTION 6

Treatment Methods in Psychiatry

6.1

The evaluation of treatments

Contents

6.1.1 The evaluation of physical treatments

Clive E. Adams

The strengths and weakness of the single trial

Strengths

New treatments, or variations of older therapies, rarely represent a revolutionary departure from what has gone before. As progress is usually made in modest steps, evaluation in prospective randomized trials is needed. These studies, comparing a new treatment with a relevant control, may be able to highlight and quantify relatively subtle but important differences in outcome.

Randomization controls for selection biases. If undertaken carefully, it should ensure that both known and unknown confounding variables, such as age, sex, and additional medications, are evenly distributed between groups. Any differences in outcome should then be due to the treatment, or the intention to give the treatment (see below). In 1991, the World Health Organization stated that the randomized controlled trial, if ethical and feasible, is the most objective means of evaluating mental health interventions.[1]

Certainly, large well-conducted trials, with participants, interventions, and outcomes recognizable to those working in health services, are potent guides to clinical practice. Nevertheless, even when such trials exist, it is important to view them alongside all other comparable evidence. Should the large study affirm the findings of smaller trials the clinician can proceed with confidence. If there is a discrepancy then debate will be generated, which should clarify important issues relating to the participants, interventions, or outcomes measured or to the methods by which the trial was conducted.[2]

Power

As numbers within a study increase so does the precision of results, enabling important but subtle differences to be detected, if they do indeed exist. Should a new treatment be considerably better than its predecessors few people would have to be randomized in order to demonstrate clearly the advantage of the innovative approach. As the advantage expected of new treatments is usually modest, reasonably large studies are often needed.

The power calculation is an important prerequisite for any randomized trial. For example, if clinical observation suggests that a new treatment can help 20 per cent more people avoid admission than the standard care, this can form the basis for a power calculation for a trial. Using a simple formula[3] the trialist can work out how many people would have to be randomized in order to have a known probability of highlighting such a difference, should one really exist. In this case, about 150 people would have to be allocated to each arm of the trial to be reasonably confident of detecting a true 20 per cent difference ($\alpha = 0.05$, $\beta = 0.8$). Most mental health trials are far too small to show up anything but very gross differences between treatments. For example, the average number of participants in schizophrenia trials is about 100 with only a slow increase over time.[4]

A single small trial should not greatly influence the clinician, but the combined results of several studies may begin to have the power to inform practice.

Biases

Randomization attempts to control for the biases that would influence treatment allocation (selection biases). Blinding at outcome attempts to control for biases that would result from participants or raters knowing which treatment had been allocated to whom (observer bias). Inadequate randomization leads to an overestimate of effect in the region of 31 to 40 per cent, and poor blinding at outcome to that of about 17 per cent.[5] Further overestimate results from the use of unpublished or modified scales, commonly seen in mental health studies, and a financial or academic investment in the therapy by the trialists.[6,7]

There are many threats to the validity of a single trial. Viewing all relevant studies, each of which was subject to different degrees of bias, should give a more balanced picture. Of course, the reader of a review should be vigilant for the systematic bias, across all trials, that may consistently sway results one particular direction.

Generalizability

Even if a study is adequately powered and undertaken with due regard for bias, a single trial may be difficult to apply to everyday practice.

(a) Participants

Most studies involve unusual participants. Frequently those eligible for trials have to give informed consent, their problems are well defined and do not involve multiple pathologies, and they are expected to tolerate the demands of a study.

(b) Interventions

Applying the results of a single study is made even more difficult because study interventions are often impractical. For example, drug trials may use rigid dose regimens impossible to apply to routine care. Psychosocial therapies tested within a trial are often of such high quality that they bear little resemblance to what an overstretched clinical service can provide.

(c) Outcomes

Measurement of outcome may also limit the value of a single trial. In a survey of 2000 schizophrenia trials, 640 different scales were used to record outcomes such as mental state, behaviour, global impression, and adverse effects.[4] Specific subspecialties within psychiatry may have an even greater propensity to create scales for trials.[8] Even within poorly powered trials, these sensitive tools may be able to detect real differences between treatments that may be statistically if not clinically significant. However, few clinicians use such scales in everyday practice and interpreting results becomes a matter of conjecture.

Trials that involve carefully defined groups of participants receiving meticulously controlled treatments and having outcomes measured on sensitive scales are called 'explanatory' studies.[9] Such trials dominate the literature, although calls for more pragmatic or 'real world' methodology are increasing[4,10] and there are now examples of this broader approach.[11,12,13,14] Currently, generalizing from the results of a single trial to day-to-day practice is inadvisable. If, however, several explanatory studies, all undertaken with constrained, but different, methodologies are giving a similar result, the clinician can feel a little more comfortable when acting on their findings.

The rogue result

Even the well-conducted generalizable trial can produce a rogue result. Currently, the acceptable level of chance is one in 20. A statistically significant result, often denoted as $p<0.05$, suggests that the finding, if the experiment was to be replicated, should occur 19 out of 20 times. One time in 20, however, a different result will appear simply because of chance. This can lead to an interesting paradox. A single trial may not provide the best evidence of how to manage people, even in the locality that the study was undertaken. The play of chance may result in an erroneous result and unless that trial is viewed in the company of all relevant evidence, clinicians will be mislead.

Time

It is inadvisable to act on the results of a single trial because of issues of power, biases, generalizability, and the possibility of chance erroneous results. There is, however, also the issue of time. Clinicians may often prefer to read the results of a single review rather than spending time assimilating information from several similar trials. Most practitioners have very little time to keep up with research relevant to their practice. Clinicians admit to having half an hour of reading time per week,[15] and much of that may not be retained.[16] Reading reviews is time efficient.

Reviews

There are two main approaches to the reviewing process—the traditional and the systematic.

The traditional review

The traditional review is often undertaken by a person well respected in the relevant field who uses knowledge and acumen, supplemented by research, to produce a synopsis of the literature. This approach still dominates the current texts, journals, and lecture tours. For example, in 1987 in four major North American medical journals, 86 per cent of review articles depended on qualitative synthesis and contained no 'methods' section whatsoever.[17] In only 6 per cent of the reviews was quantification attempted in order to support opinion and the situation did not improve much across the next decade.[18] Therefore, the clinician is left in a situation where it is difficult not to operate under a double standard. On the one hand, a large relevant trial providing objective evaluation would be desirable, but frequently, a traditional subjective review is all that is available.

Systematic reviews

The form of a systematic review encourages the introduction of basic epidemiological principles and quantification into the process of reviewing. Gene Glass, an educational psychologist, was the first to add the results of similar studies in the hope of quantifying the effects of a treatment.[19] Glass defined 'meta-analysis' as 'the statistical analysis of a large collection of analyses results from individual studies for the purpose of integrating the findings'.[20] Unsurprisingly, in the sensitive area of the psychotherapies, their first and flawed attempts in the new discipline generated controversy.[21] Critics were quick to point out that drawing conclusions from summation of very different types of therapies, undertaken by practitioners of varied experience, was likely to be inadvisable. These pioneers, who even years later are still being criticized for adding 'apples and oranges',[22] are nevertheless owed a great debt by the rest of medicine. After all, it depends on the question being asked. It is fine to mix apples and oranges, if your question is about fruit.[23]

Systematic reviews attempt to minimize bias in the identification, extraction, and summation of relevant data by applying good survey methods to the process of literature reviewing. An analogy may help. In a community survey of the prevalence of mental disorders, a researcher stands on the doorstep of the hospital and suggests that 5 per cent of the population suffers from serious mental illness. By chance, the final estimate may even be correct, but the work could not be seen as methodologically rigorous. The researcher should have written a study protocol, clearly defined an

unbiased sample of individuals to interview, and specified *a priori* the analyses to be undertaken. A systematic review should do this for a survey of a 'population' of relevant literature. Within such a review, the objectives, criteria for selection of relevant studies, search strategy, methods of study selection, data extraction, and assimilation are all made explicit.

The advantages of the systematic approach

As is suggested above, a systematic review may, by adding the results of similar studies together, at least begin to address the issue of the underpowered study, single trials of biased methodology, poor generalizability, and idiosyncratic results. Although often longer than most traditional reviews, the systematic review is still a time-efficient way to appraise research. Additional advantages are both intuitive and practical.

(a) Objectivity

Medicine remains a scientific discipline and the attempt at objective quantification must be an integral part of this approach. However, the systematic review and meta-analysis should never become a source of clinical tyranny. Individual clinicians will always have to use wisdom and judgement in their day-to-day decision-making, but to exclude objective appraisal from this process is foolhardy.

(b) Clinical empowerment

Systematic reviews can provide clear information to clinicians, policy makers, and recipients of care, and so help inform the decision-making process. For example, a systematic review of family therapy suggests that this educational, psychosocial package can help those with schizophrenia avoid or postpone relapse.[24] This finding is very much in line with the suggestions of traditional reviews.[25] However, the systematic review is able to illustrate how seven families have to undergo regular therapy, for up to a year, in order for a single relapse to be postponed. Such data, of course, mean different things to different people. Clinicians may find this an acceptable degree of effort, whereas managers of services, or even families of those with schizophrenia, may not. Although the findings may not decrease controversy, at least debate can be informed.

The quantification of trial data can sometimes provide information quite at odds with the advice of traditional reviewers. The best example comes from outside mental health care. In 1992, Antman *et al.* undertook a meta-analysis of randomized trials evaluating the care of those with acute myocardial infarction.[26] As the reviewers added trial data, they found that by 1973 enough studies existed to show clearly that thrombolytic treatment saved lives. Subsequent trials added precision to the result but did not change the finding. Antman and colleagues also showed how traditional reviews continued to fail to mention thrombolytic therapy up to 15 years after the summated trial data could have shown its value.[26] These traditional reviews recommended treatments for myocardial infarction that were positively harmful. Examples have emerged from mental health. Sometimes, traditional reviews make bold claims or recommendations which are not supported by the evidence in quantitative, systematic reviews. For example, some claims made for cognitive therapy for schizophrenia[27] go well beyond objectively summarized evidence.[28,29] Conversely, even when evidence is of high quality and readily accessible, traditional reviewers can be blind. For example, both the strengths and the limitations of new generation antipsychotic drugs have been evident for years[30,31,32] but traditional reviews and texts have encouraged uncritical enthusiasm for their use[33,34] and even guilt for failing to prescribe.[35]

(c) Gaps in research

Often a systematic review will highlight unsuspected gaps in research. The trial-base of much routine practice is not strong, and systematic reviews can help shape questions to be tested in well-planned and conducted trials.[36] Certainly, some research funders are now requiring that a systematic review be undertaken before a randomized trial is funded. This also avoids wasting resources on questions that have already been answered.

The limitations of the systematic approach

(a) Qualitative information

Systematic reviews focusing on the value of treatments given to those with mental health problems usually involves quantitative synthesis of data from randomized trials. Incorporating the great wealth of information from more qualitative approaches in an unbiased way is problematic.

(b) Trial content and quality

Systematic reviews are limited by trial content and quality. For example, it is feasible that, on average, those taking a new drug may have a statistically significant 10-point-greater decrease in a modified Brief Psychiatric Rating Scale[37] score than those taking the comparison treatment. First, this finding is difficult to put into clinically meaningful terms. Second, most scales do not provide 'interval' data. A 10-point change for someone who started with a very high score may not mean the same as the same change for a person entering the study with a lower rating. Third, more problems stem from the modification of the scale. This may well not be published and so validity is questionable. The use of such data is associated with an overestimate of effect.

Undertaking a systematic review of poor-quality trials is an important prerequisite for the design of good studies, but clinical interpretation can be problematic.[38]

(c) Rare outcomes

Randomized trials are not a powerful means of identifying rare but important outcomes. For example, large cohorts of those taking the 'atypical' antipsychotic drug, clozapine, suggest a rate of about 1 per cent for agranulocytosis, a serious adverse effect.[39] However, a systematic review of all relevant randomized trials finds a much lower incidence.[40] As the most vulnerable period for the occurrence of agranulocytosis is from weeks 6 to 18 of treatment,[39] and most studies in the systematic review were of shorter duration, the incidence was underestimated.

Trials have limited power to identify rare outcomes and, although systematic reviews may increase this power, reviews of studies of different methodologies may be needed to quantify these important effects.

(d) Limited statistical methods

The statistics used to summate data within meta-analyses are still evolving. For example, much of the continuous scale data, seen so frequently in mental health trials, is not normally distributed. How robust the commonly used methods of meta-analysis are for these non-parametric data is not clear. In addition, as mental health begins to evaluate preventive interventions then cluster randomization, where communities or institutions are randomized rather

than individuals, will become more common. The statistics for a meta-analysis of these studies is still a matter of debate.[41] Frequently, a systematic review of mental health trials must present, but not summate, relevant data.

The methods of systematic reviews

Setting the question

Clinical questions regarding the effects of treatments have three parts: the participants (who are the people of interest to the questioner?), the interventions (what are the specific treatments that are to be the focus of the review?), and the outcomes (what are the outcomes of interest to the reviewers?) . Although the reviewers may have knowledge of existing trials and their limitations, it is important that the questions set are relevant to the review's readership. If the review is to service clinicians then clinically relevant outcomes must be a priority and not necessarily those anticipated by foreknowledge of the trials. If all studies then provide data on mean change in the Brief Psychiatric Rating Scale and fail to mention the outcome of 'clinically important improvement', the review can highlight this important gap in knowledge.

Developing an answerable question. The next stage for formulating the question is to decide on the type of study that is best suited to answering the question. For questions related to the efficacy of treatments this is usually the randomized trial. At first glance, this may seem straightforward, but it is important to state *a priori* whether studies that implied, but did not state, randomization should be included. No other methodological parameter is so consistently linked with exaggerated estimations of effect than poor description of randomization.[5] Studies that describe themselves as 'prospective, double-blind, evaluative controlled trials' would be excluded from a review if the entry criteria demanded an explicit description of randomization.

Identifying studies

Studies are usually identified by searching bibliographic databases such as EMBASE, MEDLINE, or PsycINFO. Hand-searching relevant journals, conference proceedings, and references is also often undertaken.

A systematic review would be a misnomer if the researchers did not make the means of identification of studies clear and reproducible. The exact source of trials, and the search strategies, must be explicit. It is at this poin`t that the advice of an information specialist is important. The coverage of mental health journals in many bibliographic databases is poor and often limited by region or language[42] so that searching several sources is advisable. For example, the last decade China has become highly productive of mental health trials[43] but few are reported in mainstream bibliographic databases.[44]

In recent years, however, the situation for those wishing to identify all treatment trials relevant to a particular topic has become easier. The Cochrane Controlled Trials Register within the Cochrane Library,[45] is the largest and most comprehensive bibliographic database of published and unpublished randomized trials, and controlled clinical studies, in existence. For citations of trials, this specialist register has eclipsed databases such as EMBASE, Medline, and PsycINFO. Searching the Cochrane Controlled Trials Register also avoids the problem of the numerous 'false' positive citations produced by searches of unspecialized biomedical databases.

Identifying every possible study is important. Potent biases operate in this area. Trials that have statistically significant results are more likely to be published than those reporting equivocal findings,[46] and they are more likely to be published in English.[47] Systematic reviews incorporating only Anglophone published data or trials from one region are likely to produce, at best, imprecise, and at worse, overoptimistic views of efficacy.

Selecting studies

Once a search is completed, relevant studies must then be selected without bias. Reviewers usually work independently and document the outcome of all disputed decisions. Some feel that those selecting the trials should be blind to the study's author, source (usually a journal), and the institution where the trial was undertaken. All have potential to bias the study selection. However, such blinding would often involve prohibitive effort. In any event, a systematic review should make explicit the degree of effort made to avoid selection bias at this crucial stage.

Quality assessment of trials

Once studies are identified and prespecified entry criteria met, a last set of quality criteria may be applied. Scales are available, but essentially they rate selection and observation bias (see above). The description of concealment of allocation is central, as this methodological parameter has consistently been shown to be linked with an estimate of effect. If this is poorly reported, the trial is likely to overestimate the effect of the experimental intervention.[5] For trials that describe allocation with nothing more than 'randomized', this single parameter may not be a sensitive measure of quality. A scale addressing both selection and observation bias by rating the description of randomization, blinding, and reasons for people withdrawing, may be more appropriate.[48]

A systematic review should prestate the level of quality that is acceptable, or, at the very least, how the data from poor-quality studies are to be managed.

Data extraction

Reliable data extraction is important. Just as studies must be selected with due regard for the inclusion of bias, so data must be extracted carefully and reproducibly. Often reviewers ensure maximum reliability by organizing double data extraction by an independent reviewer.

Data management

(a) *A priori* primary analysis

As with any quantitative research, a systematic review will generate the potential for multiple analyses. As one in 20 will be statistically significant by chance, it is important to state *a priori* the primary analyses to be undertaken. Although multiple secondary analysis are often undertaken, these are only hypothesis-generating as data have been multiply tested.

(b) Unacceptable loss to follow-up

In every study, there must be a certain attrition that renders data meaningless. For example, in a trial of tacrine for those with Alzheimer's disease 68 per cent of people taking the experimental compound were withdrawn or lost to follow-up.[49] Drawing conclusions from the data provided by the 32 per cent of 'completers' is problematic as selection bias, originally addressed by

randomization, is likely to be great. Trial attrition may not be immediately apparent from first glance. For example, a meta-analysis of studies comparing the antipsychotic quetiapine with chlorpromazine and haloperidol for schizophrenia shows considerable loss to follow-up at only a few weeks.[50] The last observation of those leaving was carried forward to the results, so that data presented in the trials were on the numbers originally randomized. The trialists made an assumption that data collected just before leaving the study would reflect the situation at the end of the trial. These assumptions by the trialists may or may not be justified[51] but reviewers must also make judgements. It is crucial to make these important decisions explicit and to make them before seeing the data.

The limit at which data become meaningless may differ depending on the question addressed. For example, in the situation of trialling a new oral drug for schizophrenia, clearly a loss of nearly 60 per cent of people at 6 weeks is clinically untenable. The reviewer may judge that the unfortunate clinician may lose up to 30 per cent of people by 6 weeks but that any greater loss would reflect more than misfortune and render data of little use. In different circumstances, such as the acute care of very disturbed people in closed wards, the loss of even 10 per cent of participants could be seen as a threat to the value of the data presented.

(c) Intention-to-treat analysis

Interventions are not randomized in trials — it is the *intention* to give treatments that is randomly allocated. Once people are lost to follow-up, the property of the randomization to distribute known and unknown confounding variables is under threat. The randomization has, in effect, been broken. The real threat of an introduction of selection bias has led to the phrase — once randomized, always analyze.[3]

Once a limit to trial attrition has been set, reviewers must, before seeing trial data, exercise more judgement in what outcome is to be attributed to those who were lost. It is impossible to avoid assumptions, but these should be based on common sense if not evidence. For example, when presenting data for the outcome of 'clinically improved', reviewers could assume, unless contrary information is provided in the trials, that those who left early did not have an important recovery. If good quality sources of information are available, this assumption can become more evidence-based. Perhaps an exemplary trial within a systematic review managed 100 per cent follow-up even on those who left the study early. If this trial found that 90 per cent of those who had not complied with the study protocol were not 'clinically improved', it would provide a rationale for applying this figure to the other trials in the meta-analysis. Unless individual patient data are available, this process is impossible for continuous outcomes and only 'completer' data must be presented.

(d) Continuous data

Data on continuous outcomes are frequently skewed, with the mean not being the centre of the distribution. The statistics for meta-analysis are thought to be able to cope with some skew, but were formulated for parametric normally distributed data. Reviewers may wish to build in simple rules to avoid the potential pitfall of applying parametric tests to very skewed data. For example, in scale data where a mean endpoint score is provided with a standard deviation, when the latter is multiplied by 2 and is then greater than the mean, data could be stated to be too skewed to summate.[52] This rule cannot be applied for scale data reporting change, rather than endpoint, scores.

A wide range of rating scales are available to measure outcomes in mental health trials. These scales vary in quality and many are poorly validated. It is generally accepted that measuring instruments should have the properties of reliability (the extent to which a test effectively measures anything at all) and validity (the extent to which a test measures that which it is supposed to measure). Before publication of an instrument, most scientific journals insist that both reliability and validity be demonstrated to the satisfaction of referees. Reviewers may well decide, as a minimum standard, to exclude data from unpublished rating scales.[38]

(e) Individual patient data

Most mental health meta-analyses are of aggregate data from published reports. Other specialities have set a 'gold standard' for systematic reviews by acquiring, checking, and reanalyzing each person's data from the original trialists.[53] Collecting individual patient data allows reviewers to undertake time-to-event analyses and subgroup analyses, to ensure the quality of the randomization and data through detailed checking and correction of errors by communication with triallists, and finally to update follow-up information through patient record systems (such as mortality registers).

Limited empirical evidence exists for some of the advantages of individual patient data reviews over other types of review. The former does help to control publication bias, to ensure use of the intention-to-treat principle in the analysis, and to obtain a fuller picture of the effects of different treatments over time. Undertaking individual patient data reviews requires considerable additional skills, time, and effort on the part of the reviewers when compared to meta-analyses of published aggregate data.[54]

Statistics

(a) Inappropriate meta-analyses

Systematic reviews may, or may not, contain meta-analyses. Where participants, interventions, or outcomes are clearly too different to summate, reviewers must resist the temptation to use powerful statistics on inappropriate data.[41]

(b) Summary measures

Much has been written about the statistics for meta-analysis,[55] but if their meaning is not conveyed to the user of the review, they are of little value. Summary measures such as odds ratios or relative risk are frequently employed for dichotomous outcomes and weighted and standard mean difference for continuous data. Where continuous data are presented from different scales measuring similar phenomena then standard mean difference is often calculated and called the effect size. This estimate has statistical integrity, but is even more problematic to interpret clinically than weighted mean difference. In each of these summary measures, an individual trial contributes to the final statistic, inversely proportional to the precision of its result.

Currently, in this new discipline, statistics for meta-analysis are powerful and evolving. In recent years, there has been a much better understanding of how best to summate data from cluster randomized trials and crossover studies. Techniques for these types of analyses are now widely accessible. Further statistical flexibility is available through use of generic inverse variance. This statistical

technique facilitates analyses of properly analyzed cross-over trials, cluster randomized trials and non-randomized studies, as well as outcome data that are ordinal, time-to-event or rates. Better methods for managing non-parametric data can confidently be expected in the next few years.[41]

(c) Sensitivity analyses

This is where an analysis is used to determine how sensitive the results of the review are to changes in how it was undertaken. For example, the reviewers may state, again *a priori*, that they wish to compare the size of effect of industry-sponsored trials versus those undertaken independently of the manufacturer of the experimental drug.[7] The sensitivity of the final result to adding and subtracting sets of trials is then tested. Sensitivity analyses can be proposed on many variables, such as severity of illness, age of participant, means of diagnosis, subtype of intervention, and quality of trial. This can easily lead to the problems of multiple testing, although for meta-analyses of published data the quality and extent of trial reporting severely restricts the numbers of sensitivity analyses that are possible.

(d) Heterogeneity

In systematic reviews heterogeneity refers to variability or differences between studies' estimates of effects and is a function of clinical and/or methodological diversity among the studies. Despite rigorous definition and application of inclusion criteria, the trials eventually selected may not be homogeneous enough to summate. Heterogeneity should be considered, sought, and measured, and, if present, investigated. Statistical tests of heterogeneity are used to assess whether the observed variability in study results (effect sizes) is greater than that expected to occur by chance. These tests, however, have low statistical power and careful inspection of results for outlying findings is just as valuable. Recently, partly because it is more intuitive than previous measures, the I^2 has become widely used. This quantifies inconsistency across studies, moves the focus away from testing whether heterogeneity is present to assessing whether the heterogeneity, that many argue is inevitable, will impact on the meta-analysis. A value greater than 50 per cent is often considered to indicate substantial heterogeneity.[41]

Heterogeneity can be caused by various factors, and its presence generates debate about differences in study design (methodological diversity) and differences between studies in key characteristics of the participants, interventions, or outcome measures (clinical diversity). Heterogeneity can be explored by undertaking subgroup analyses or employing meta-regression. In subgroup analyses, the results of one group of studies with a characteristic thought to be causing heterogeneity is compared to those of another group of trials without that characteristic. For example, trials of a new interpersonal therapy for depression may be heterogeneous. The reviewers may have stated, *a priori*, that the new therapy, targeted at young people, may not be effective for those over 30 years of age. Overall, although the trials may have heterogeneous results, if the two subgroups of studies involving people over and under the age of 30 have homogeneous results, it is feasible that the original hypothesis of the reviews regarding age was correct. Meta-regression, is an extension to subgroup analyses that allows the effect of variables, even several simultaneously, to be investigated. Again these variables should be pre-stated and this technique necessitates much high quality data to be meaningful. Even then, the regression is essentially an observational study with all the dangers of unexplained or residual confounding. In mental health reviews, investigation of heterogeneity often generates valuable debate, rather than providing definitive answers.

(e) Reporting bias

There are several ways to assess whether reporting (publication) bias (see above) is operating within a review. The reviews may use a funnel plot technique[56] where the results of a trial are plotted against its size. Large studies with any result, positive or negative, tend to be published. Small positive studies are also usually easily identified, but it quickly becomes apparent if small 'negative' studies have not been found. This could be due to a variety of reasons, but one of which is a selective reporting of positive outcomes.

Sources of systematic reviews

Journals

Systematic reviews are increasingly seen in major journals. These can be identified by simple or comprehensive methodology-specific searches of bibliographic databases such as EMBASE, Medline, or PsycINFO. Examples of these searches can be readily identified[57] but need to be updated periodically to reflect changes in notation or indexing. These search phrases would then be linked to a subject-specific phrase such as 'depression' or 'cognitive therapy', by the word 'and' to limit and focus the number of identified citations.

DARE database

The Database of Abstracts of Reviews of Effectiveness (DARE) is a specialist database containing thousands of citations of systematic reviews of health care. These reviews are assessed according to explicit criteria, and structured abstracts describing methodology and results, and commenting on quality and clinical implications, are also included.

Cochrane reviews

The Cochrane Database of Systematic Reviews is an electronic publication increasingly favoured by clinicians, researchers and those compiling guidelines. It contains the full text and data of reviews undertaken to the most rigorous systematic standards[58] and is updated quarterly. There is often a considerable lag time for traditional journal publication and this can result in the publication of good-quality but misleading systematic reviews. For example, an important systematic review on the effects of family intervention for schizophrenia was published in August 1994.[59] Just 2 weeks later, the same authors re-summated relevant data in an electronic version and the results were much less favourable than had been previously reported.[60] In the considerable period between completion of the paper version and its publication, less favourable trials appeared. Further updates of this review suggest that the trend is continuing. By using electronic publishing, the Cochrane Database of Systematic Reviews allows trends over time to be highlighted and reviews to be maintained.

The Cochrane Library is a collection of databases supplying high-quality evidence to inform all those interested in the evaluation of health care.[45] It is published quarterly on CD-ROM and the Internet, and includes the Cochrane Database of Systematic Reviews, the source of many of the reviews quoted in this

chapter, the DARE abstracts, and The Cochrane Controlled Trials Register.

Summary publications

There are now several periodically updated, systematically compiled, publications designed primarily for busy clinicians. For example, one such publication, Clinical Evidence,[61] now has a comprehensive mental health subset, and is available in short or very short versions, providing clinical 'bottom lines' of relevant systematic reviews or randomized trials.

Further information

The Cochrane Handbook - http://www.cochrane-handbook.org/ contains up-to-date methodological details of systematic reviews and meta-analysis

Clinical Evidence - http://clinicalevidence.bmj.com/ceweb/index.jsp an on-line, regularly updated compendium of systematic evidence summaries

References

1. WHO Scientific Group on Treatment of Psychiatric Disorders (1991). Evaluation of methods for the treatment of mental disorders. *World Health Organization Technical Report Series*, **812**, 1–75.
2. Egger, M. and Davey-Smith, G. (1995). Misleading meta-analysis. *British Medical Journal*, **311**, 753–4.
3. Pocock, S. (1989) *Clinical trials: a practical approach*. Wiley, Chichester.
4. Thornley, B. and Adams, C. (1998). Content and quality of 2000 randomised controlled trials in schizophrenia over 50 years. *British Medical Journal*, 1181–4.
5. Juni, P., Altman, D. and Egger, M. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. *British Medical Journal*, **323**, 42–6.
6. Heres, S., Davis, J., Maino, K., *et al.* (2006). Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of headto- head comparison studies of second-generation antipsychotics. *American Journal of Psychiatry*, **163**, 185–94.
7. Montgomery, J., Byerly, M., Carmody, T., *et al.* (2004). An analysis of the effect of funding source in randomized clinical trials of second generation antipsychotics for the treatment of schizophrenia. *Controlled Clinical Trials*, **25**, 598–612.
8. Cure, S., Chua, W., Duggan, L. *et al.* (2005). Randomised controlled trials relevant to aggressive and violent people, 1955-2000: a survey. *British Journal of Psychiatry*, **186**, 185–9.
9. Roland, M. and Torgerson, D. (1998). What are pragmatic trials? *British Medical Journal,* **316**, 285.
10. Simon, G., Wagner, E. and Vonkorff, M. (1995). Cost-effectiveness comparisons using «real world» randomized trials: the case of new antidepressant drugs. *Journal of Clinical Epidemiology*, **48**, 363–73.
11. Lieberman, J., Stroup, T., McEvoy, J., *et al.* (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*, **353**, 1209–23.
12. Lewis, S., Davies, L., Jones, P., *et al.* (2006). Randomised controlled trials of conventional antipsychotic versus new atypical drugs, and new atypical drugs versus clozapine, in people with schizophrenia responding poorly to, or intolerant of, current drug treatment. *Health Technology Assessment*, **10**, iii-iv, ix-xi, 1–165.
13. TREC Collaborative Group (2003). Rapid tranquillisation for agitated patients in emergency psychiatric rooms: a randomised trial of midazolam versus haloperidol plus promethazine. *British Medical Journal*, **327**, 708–13.
14. Alexander, J., Tharyan, P., Adams, C., *et al.* (2004). Rapid tranquillisation of violent or agitated patients in a psychiatric emergency setting. Pragmatic randomised trial of intramuscular lorazepam v. haloperidol plus promethazine. *British Journal of Psychiatry*, **185**, 63–9.
15. Sackett, D., Richardson, W., Rosenberg, W. *et al.* (1997). *Evidence based medicine: how to practice and teach EBM*. Churchill Livingstone, London.
16. Kellett, C., Hart, A., Price, C., *et al.* (1996). Poor recall performance of journal-browsing doctors. *Lancet*, **348**, 479.
17. Mulrow, C. (1987). The medical review article: state of the science. *Annals of Internal Medicine*, **106**, 485–8.
18. McAlister, F., Clark, H., van, W.C., *et al.* (1999). The medical review article revisited: has the science improved? *Annals of Internal Medicine*, **131**, 947–51.
19. Smith, M. and Glass, G. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, **32**, 752–60.
20. Glass, G. (1976). Primary, secondary and meta-analysis of research. *Education Research*, 3–8.
21. Eysenck, H. (1978). An exercise in mega-silliness. *American Psychologist*, **33**, 517.
22. Eysenck, H. (1994). Meta-analysis and its problems. *British Medical Journal*, **309**, 789–92.
23. Higgins, J. and Green, S. (2005). *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005]*, Chichester, UK, John Wiley & Sons, Ltd.
24. Pharoah, F., Mari, J., Rathbone, J. *et al.* (2006). Family intervention for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000088.
25. Leff, J. (1994). Working with the families of schizophrenic patients. *British Journal of Psychiatry. Supplement*, 71–6.
26. Antman, E., Lau, J., Kupelnick, B., *et al.* (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *Journal of American Medical Association*, **268**, 240–8.
27. Kingdon, D. (2006). Psychological and social interventions for schizophrenia. *British Medical Journal*, **333**, 212–3.
28. McKenna, P. (2006). What works in schizophrenia: cognitive behaviour therapy is not effective. *British Medical Journal*, **333**, 353.
29. Jones, C., Cormac, I., Silveira, D.M.N.J. *et al.* (2004). Cognitive behaviour therapy for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000524.
30. Srisurapanont, M., Disayavanish, C. and Taimkaew, K. (2000). Quetiapine for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000967.
31. Duggan, L., Fenton, M., Dardennes, R., *et al.* (2000). Olanzapine for schizophrenia. *Cochrane Database of Systematic Reviews*, CD001359.
32. Kennedy, E., Song, F., Hunter, R., *et al.* (2000). Risperidone versus typical antipsychotic medication for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000440.
33. Adams, C. and Jayaram, M. (2007). Do findings from new trials for schizophrenia fit with existing evidence: not duped ……. just beguiled? *Epidemiologia e Psichiatria Sociale*, **16**, 199–202.
34. Vedantam, S. (Oct 3, 2006). In antipsychotics, newer isn't better – drug find shocks researchers. *Washington Post* A01.
35. Adams, C., Tharyan, P., Coutinho, E. *et al.* (2006). The schizophrenia drug-treatment paradox: pharmacological treatment based on best possible evidence may be hardest to practise in high-income countries. *British Journal of Psychiatry*, **189**, 391–2.
36. Brown, P., Brunnhuber, K., Chalkidou, K., *et al.* (2006). How to formulate research recommendations. *British Medical Journal*, **333**, 804–6.
37. Overall, J. and Gorham, D. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports*, **10**, 799–812.
38. Marshall, M., Lockwood, A., Bradley, C., *et al.* (2000). Unpublished rating scales: a major source of bias in randomised controlled

trials of treatments for schizophrenia. *British Journal of Psychiatry*, **176**, 249–52.
39. Atkin, K., Kendall, F., Gould, D., *et al.* (1996). Neutropenia and agranulocytosis in patients receiving clozapine in the UK and Ireland. *British Journal of Psychiatry*, **169**, 483–8.
40. Wahlbeck, K., Cheine, M. and Essali, M. (2000). Clozapine versus typical neuroleptic medication for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000059.
41. Deeks, J., Higgins, J. and Altman, D. (2005). Analysing and presenting results. In: Higgins, J. and Green, S., (Eds.) *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005] Section 8. In: The Cochrane Library*, John Wiley & Sons, Ltd., Chichester, UK.
42. McDonald, S., Taylor, L. and Adams, C. (1999). Searching the right database. A comparison of four databases for psychiatry journals. *Health Libraries Review*, **16**, 151–6.
43. Chakrabarti, A., Adams, C., Rathbone, J., *et al.* (2007). Schizophrenia trials in China: a survey. *Acta Psychiatrica Scandinavicia*, (in press).
44. Xia, J., Wright, J. and Adams, C. (2007). Five large Chinese biomedical bibliographic databases: accessibility and coverage. *Health Infomation Libraries Journal*, (in press).
45. Cochrane Collaboration (2007). *The Cochrane Library*, (3rd edn.). John Wiley & Sons, Ltd., Chichester, UK.
46. Easterbrook, P., Berlin, J., Gopalan, R. *et al.* (1991). Publication bias in clinical research. *Lancet*, **337**, 867–72.
47. Egger, M., Zellweger-Zahner, T., Schneider, M., *et al.* (1997). Language bias in randomised controlled trials published in English and German. *Lancet*, **350**, 326–9.
48. Jadad, A., Moore, R., Carroll, D., *et al.* (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials*, **17**, 1–12.
49. Knapp, M., Knopman, D., Solomon, P., *et al.* (1994). A 30-week randomized controlled trial of high-dose tacrine in patients with Alzheimer's disease. The Tacrine Study Group. *Journal of American Medical Association*, **271**, 985–91.
50. Srisurapanont, M., Maneeton, B. and Maneeton, N. (2004). Quetiapine for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000967.
51. Leucht, S., Engel, R., Bauml, J. (2007). Is the superior efficacy of new generation antipsychotics an artifact of LOCF? *Schizophrenia Bulletin*, **33**, 183–91.
52. Altman, D. and Bland, J. (1996). Detecting skewness from summary information. *British Medical Journal*, **313**, 1200.
53. Early Breast Cancer Trialist's Collaborative Group (1992). Systematic treatment of early breast cancer by hormonal, cytoxic, or immune therapy: 133 randomised trials involving 31 000 recurrences and 24 000 deaths among 75 000 women. *Lancet*, **1-15**, 71–85.
54. Stewart, L. and Clarke, M. (1995). Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. *Statistics in Medicine*, **14**, 2057–79.
55. Petitti, D. (1994). *Meta-analysis, decision analysis and cost-effectiveness analysis. Methods for quantitative synthesis in medicine*. Oxford University Press, Oxford.
56. Egger, M., Davey-Smith, G., Schneider, M. *et al.* (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, **315**, 629–34.
57. Shojania, K. and Bero, L. (2001). Taking advantage of the explosion of systematic reviews: an efficient MEDLINE search strategy. *Effective Clinical Practice*, **4**, 157–62.
58. Jadad, A., Cook, D., Jones, A., *et al.* (1998). Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *Journal of American Medical Association*, **280**, 278–80.
59. Mari, J. and Streiner, D. (1994). An overview of family interventions and relapse on schizophrenia: meta-analysis of research findings. *Psychological Medicine*, **24**, 565–78.
60. Mari, J. and, S.D. Family intervention for schizophrenia. *The Cochrane Library*.
61. Tovey, D. (2007). *BMJ Clinical Evidence*, BMJ Publishing Group Ltd., London.

6.1.2 The evaluation of psychological treatment

Paul Crits-Christoph and
Mary Beth Connolly Gibbons

Introduction

Psychotherapy continues to be a widely practised treatment for psychiatric disorders and other problems in living. Since publication in 1952 of the well-known article by Hans Eysenck,[1] in which he claimed that there was no evidence that psychotherapy was effective, there has been an accelerating literature concerned with methodologies for evaluating psychotherapy, as well as specific studies demonstrating the efficacy, or lack thereof, of various psychotherapies. In more recent years, pressures from the government agencies and insurance companies that bear much of the cost of mental health treatments have added to the call for accountability regarding psychotherapeutic treatment.

Despite a vast literature of over 1000 outcome studies of the effects of psychotherapy, questions remain about the role of psychotherapy as a treatment for mental disorders. Extensive meta-analytical reviews of the psychotherapy outcome literature provided evidence that, generally speaking, psychotherapy appears to be efficacious.[2] While encouraging, this information was not particularly useful. As with any medical problem or disorder, the relevant public health clinical question is whether a treatment is beneficial for the presenting problem or psychiatric disorder for which help is sought. Along these lines, a number of efforts have been made at summarizing the results of the psychotherapy outcome literature in terms of what works for different disorders or problems.[3, 4] For example, these efforts have arrived at conclusions such as 'cognitive therapy is efficacious in the treatment of major depressive disorder'.

The simplicity and clinical appeal of such conclusions, about which psychotherapy treatments work for which patient problems, belies a host of more complex issues regarding how one evaluates psychotherapy and makes a decision about whether treatment 'works' or not. Other treatments within psychiatry, such as pharmacotherapy, lend themselves to rather straightforward designs (namely placebo-controlled randomized clinical trials) that permit clear inferences about the efficacy of a treatment approach. In contrast, research on psychotherapy as a verbal interchange between two or more participants does not have the luxury of such straightforward pharmacotherapy research designs. Instead, psychotherapy outcome research is characterized by the use of a variety of research designs and methods that, while often not without limitations to strong scientific inferences about treatment efficacy, can provide incremental scientific advance in the understanding of the usefulness of psychotherapeutic treatments. The aim of the current chapter is to provide an overview of approaches to the evaluation

of psychological treatments. We begin with a discussion of specific research designs employed in psychotherapy outcome research, with a discussion of some of the broad issues that currently guide the selection among these different experimental designs. This is followed by a selective review of assessment strategies for outcome evaluation, with discussion of examples of instruments.

Issues in planning research evaluating psychotherapy

A number of other sources provide a detailed discussion of issues involved in planning a study on psychotherapy, as well as explication of various research designs. In particular, our presentation draws heavily from Kazdin,[5] supplemented with writings that illustrate more recent trends in both design and methodology.

There are, of course, a wide range of decisions to be made in designing an evaluation study of psychotherapy. These decisions affect the choice of patients, therapists, control groups, data analytical strategies, etc. Table 6.1.2.1 presents a list of the types of questions that need to be asked in designing or evaluating a study of psychotherapy outcome. A discussion of some of the key methodological issues that cut across many of the questions raised in Table 6.1.2.1 follows.

Internal versus external validity

An initial important decision in planning an evaluation of psychotherapy outcome, or any intervention, is the relative emphasis on internal versus external validity of the inferences from the investigation. Internal validity refers to the extent to which inferences can be attributed to the intervention *per se*, as opposed to other factors. In order to maximize internal validity, the investigator attempts to control as many of the extraneous factors as possible through a variety of procedures including, among others, random assignment, the use of control groups, assessing subjects in the same ways and at the same point in time, and careful selection of a relatively homogeneous subject sample. With as many factors as possible held constant across treatment groups except for the nature of the intervention, an outcome difference detected between the experimental and control group(s) can be attributed to the intervention, rather than other factors.

External validity, in contrast, refers to the extent to which the results of a study can be generalized to other subjects, settings, treatment durations, and treatment providers other than those used within the specific study. In regard to the evaluation of psychotherapy outcomes, external validity is often invoked to raise the question of whether study results pertain to the 'real world' in which psychotherapy is practiced—the diverse set of patients, therapists, and settings occurring in the community that may be quite different from the conditions of an investigation conducted in a research setting.

Clearly, both internal and external validity are important, but it is difficult to maximize each within the context of the same study. Studies of homogeneous patient samples, for example, may have high internal validity, but generalize poorly to the mix of heterogeneous patients seen in clinical practice. The relative merits of studies with high internal versus external validity have been a source of ongoing debate among psychotherapy outcome researchers.

Different research designs are more or less appropriate depending upon the scientific question of interest. For example, the process of developing and testing new treatments generally proceeds stepwise, beginning with individual case reports and then progressing to an 'open-label' (a term derived from pharmacotherapy research) trial involving the application of a single treatment to a relatively small group of patients. Following an open-label trial, a promising treatment would then be tested within the context of a controlled, efficacy trial. In this efficacy trial, the treatment would be tested under ideal circumstances (for instance, by highly trained clinicians). If an effect is found in the controlled efficacy trial, a controlled, effectiveness trial is the next step. In this effectiveness trial, the treatment would now be tested under more 'real world' conditions. This line of research is oriented towards understanding whether the treatment *per se* is responsible for change (efficacy trial) and whether the effect generalizes (effectiveness trial).

Naturalistic studies represent an alternative type of effectiveness trial in which the scientific question is usually not a focus of type of treatment. Instead, such studies might examine the relationship of patient characteristics, therapist factors, or length (dose) of treatment to outcome.

Selection criteria for psychotherapy outcome studies

The choice of selection criteria for a psychotherapy outcome study depends, of course, on the nature of the research question to be asked. From a public health perspective, samples are usually chosen based upon the presence of a discrete disorder or problem that has significance to society. The selection of the target disorder, however, is only the beginning of the selection process. For studies of DSM Axis I non-psychotic disorders, it is typical that other major psychotic disorders such as schizophrenia and bipolar disorder are excluded from the study. However, there is wide variability across research studies in the extent to which other Axis I and Axis II disorders are included in a study or not.

This aspect of selection criteria relates primarily to the internal versus external validity distinction discussed above. Studies that emphasize internal validity will probably exclude many comorbid diagnoses, while studies that maximize external validity will tend to be more inclusive. As the comorbidities among Axis I diagnoses can be high, the impact on the nature of the patient sample selected can be considerable.

Naturalistic studies that focus on psychotherapy *per se*, rather than public health concerns, are oriented towards external validity and typically do not have restrictive selection criteria. For these studies, the question is 'how effective is psychotherapy for the types of patients that end up in psychotherapeutic treatment in the community?' Thus, few, if any, selection criteria are specified.

One particular selection problem that affects any type of psychotherapy outcome study is whether or not patients currently treated with psychotropic medication are included in the evaluation study. Once again, from the point of view of internal validity—attempting to attribute the treatment outcome to the psychotherapy treatment *per se*—patients on concurrent medication treatment are usually excluded. In contrast, external validity concerns would lead to the inclusion of patients on medications, since increasing numbers of patients in the community with anxiety and affective

Table 6.1.2.1 Selected questions to raise in planning a study of psychotherapy

Sample characteristics
1 Who are the subjects and how many of them are there in this study?
2 Why was this sample selected in light of the research goals?
3 How was this sample obtained, recruited, and selected?
4 What are the subject and demographic characteristics of the sample (e.g. sex, age, ethnicity, race, socio-economic status)?
5 What, if any, inclusion and exclusion criteria were invoked (i.e. selection rules to obtain participants)?
6 How many of those subjects eligible or recruited actually were selected and participated in the study?
7 With regard to clinical dysfunction or subject and demographic characteristics, is this a relatively homogeneous or heterogeneous sample?
Design
1 How were subjects assigned to groups or conditions?
2 How many groups were included in the design?
3 How are the groups similar and different in how they are treated in the study?
4 Why are these groups critical for addressing the questions of interest?
Procedures
1 Where was the study conducted (setting)?
2 What measures, materials, equipment, and/or apparatus were used in the study?
3 What is the chronological sequence of events to which subjects were exposed?
4 What intervals elapsed between different aspects of the study (assessment, treatment, follow-up)?
5 What variation in administration of conditions emerged over the course of the study that may introduce variation within and between conditions?
6 What procedural checks were completed to avert potential sources of bias in implementing the manipulation and assessment of dependent measures?
7 What checks were made to ensure that the conditions were carried out as intended?
8 What other information does the reader need to know to understand how subjects were treated and what conditions were provided?
Therapists
1 Who are the therapists, and why are these individuals selected?
2 Can the influence of the therapist be evaluated in the design as a 'factor' (as in a factorial design) or can therapist efforts be evaluated within a condition?
3 Are the therapists adequately trained? By what criteria?
4 Can the quantity and quality of their training and implementation of treatment be measured?
Treatment
1 What characteristics of the clinical problem or cases make this particular treatment a reasonable approach?
2 Does the version of treatment represent the treatment as it is usually carried out?
3 Does the investigation provide a strong test of treatment? On what basis has one decided that this is a strong test?
4 Has treatment been specified in manual form or have explicit guidelines been provided?
5 Has the treatment been carried out as intended? (Integrity is examined during the study but evaluated after it is completed.)
6 Can the degree of adherence of therapists to the treatment manual be codified?
7 What defines a completed case (e.g. completion of so many sessions)?
Assessment
1 If specific processes in the clients or their interpersonal environment are hypothesized to change with treatment, are these to be assessed?
2 If therapy is having the intended effect on these processes, how would performance be evident on the measure? How would groups differ on this measure?
3 Are there additional processes in therapy that are essential or facilitative to this treatment, and are these being assessed?
4 Does the outcome assessment battery include a diverse range of measures to reflect different perspectives, methods, and domains of functioning?
5 What data can be brought to bear regarding pertinent types of reliability and validity for these measures?
6 Are treatment effects evident in measures of daily functioning (e.g. work, social activities)?
7 Are outcomes being assessed at different times after treatment?
Data evaluation
1 What are the primary measures and data upon which the predictions depend?
2 What statistical analyses are to be used and how specifically do these address the original hypotheses and purposes?
3 Are the assumptions of the data analyses met?
4 What is the likely effect size that will be found based on other treatment studies or meta-analyses?
5 Given the likely effect size, how large a sample is needed to provide a strong powerful test of treatment (e.g. power ≥ 0.80)?
6 Are there subdivisions of the sample that will be made to reduce the power of tests of interest to the investigator?
7 What is the likely rate of attrition over the course of treatment, and post-treatment and follow-up assessments?
8 With the anticipated loss of cases, is the test likely to be sufficiently powerful to demonstrate differences between groups if all cases complete treatment?
9 If multiple tests are used, what means are provided to control error rates?
10 Prior to the experimental conditions, were groups similar on variables that might otherwise explain the results (e.g. diagnosis, age)?
11 Are data missing due to incomplete measures (not filled out completely by the subject(s) or loss of subjects)? If so, how are these handled in the data analyses?
12 Will the clinical significance of client improvement be evaluated and if so by what method(s)?
13 Are there ancillary analyses that might further inform the primary analyses or exploratory analyses that might stimulate further work?

disorders are receiving psychotropic medications for their problems. Often, a compromise is struck: patients on medications are eligible for the psychotherapy study as long as they (and their prescribing doctor) agree to maintain a stable dosage of the medication for the duration of the psychotherapy study.

Treatment standardization

Psychotherapy efficacy research, like pharmacotherapy research, requires that the treatment be standardized. Such standardization serves two related purposes. First, from a clinical point of view, it is necessary that the treatment be clearly specified, so that any conclusions about differential treatment efficacy can be translated into clear treatment recommendations. From a research point of view, treatment standardization allows studies to be replicated. In addition, by making the delivery of a treatment more standardized, differences between therapists and the statistical problems that result from the non-independence that 'therapist effects' introduce can be avoided.[6]

Standardization of pharmacological interventions is relatively straightforward—a per-day dosage (or range of dosages) is set in advance. But for psychotherapy, how can something so complex as patient–therapist dialogue be standardized? The central ingredient in standardization of a psychosocial treatment is a treatment manual. A psychotherapy manual describes the treatment in detail, with case examples and instruments for psychotherapists. Some treatment manuals, particularly those coming from a cognitive behavioural perspective, present a highly systematized step-by-step programme which therapists follow over the course of therapy. The relative success of treatment manuals in standardizing psychotherapy has been supported by a meta-analysis,[7] which documented that studies employing treatment manuals had fewer outcome differences between therapists compared with studies that did not employ treatment manuals. Thus, when a treatment manual is used, therapists appear to produce relatively more uniform outcomes. In contrast, when no treatment manual is used, therapists differ considerably in their typical outcomes with patients; suggesting that different therapists are likely to be conducting sessions in discrepant ways, with some therapists producing more favourable outcomes and other therapists producing less favourable outcomes.

Treatment standardization, however, does not simply translate to the use of a treatment manual. A variety of steps are needed to ensure that therapists are delivering the intended treatment (Table 6.1.2.2), including: the careful selection of therapists; training of therapists in the intended modality using a treatment manual; certification of therapists based upon their adherence to the treatment model during training; and continuing adherence and competence monitoring of therapists during a clinical trial.

Concerns have been raised about the 'treatment manual' concept applied to less directive treatments such as psychodynamic therapy. The belief is that session-by-session manuals would remove the essence of good psychotherapy, and good dynamic therapy in particular, by making treatment artificially rigid and taking away the necessary clinical flexibility and creativity. Psychodynamic treatment manuals, however, are perhaps better described as 'guides', which specify the principles of treatment but do not overly constrain the necessary clinical flexibility and creativity. The flexibility of treatment is fully retained through the principle of tailoring the treatment intervention to the specific idiosyncratic issues that are salient for each patient. The actual learning of the practice of treatment is accomplished through supervision in the application of the treatment manual. Because dynamic treatment manuals are less like 'cookbooks', there may be a greater reliance on the supervision process compared with perhaps more straightforward behavioural treatments.

Table 6.1.2.2 Steps involved in the standardization of psychotherapy for outcome research

Selection of therapists
Training of therapists using a treatment manual
Certification of therapists based upon adherence to the treatment model
Continued assessment of therapist adherence and competence during a clinical trial

Research designs

Unlike pharmacological research where a single form of research design (placebo-controlled study) dominates the literature, psychotherapy researchers have employed a host of different research designs to understand the effects of psychotherapy. Some of the more common designs are listed in Table 6.1.2.3 and are explicated in the next section.

Single-case designs

Clinical evaluation of the effects of psychotherapy dates back to Freud's descriptions of individual cases in treatment. However, methods to systematically examine the effects of interventions with individual patients were developed by behaviour therapists.[8] These single-case experiments rely on comparing patient responses to differing experimental conditions over time. Typically, such single-case studies begin with an extended baseline period where patient behaviours or symptoms are recorded without any intervention. Then, different intervention phases are introduced, usually followed by more baseline (no intervention) assessment phases.

While experimental, single-case designs lend themselves well to the investigation of behavioural treatments that include a focus on immediate overt behaviour, such designs have rarely been employed with other verbal psychotherapies that emphasize longer term processes such as patient psychological growth and functioning. The generalizability of findings from single-case research is another limitation to this form of research.

Table 6.1.2.3 Common research designs for the evaluation of psychotherapy

Single-case designs
Randomized controlled trials with non-specific or psychological 'placebo' control
Comparative designs
Dismantling or additive designs
Comparisons with medication and pill placebos
Naturalistic designs

Randomized controlled trials

Random assignment of subjects to treatment and control conditions is generally viewed as the preferred method of evaluating the effects of interventions in psychiatry, and in medicine in general. However, perhaps the single most vexing problem in research of psychotherapy outcome is the design of control groups. In pharmacotherapy research, a pill placebo (with 'double blinding', i.e. the patient and the doctor are unaware of treatment assignment) serves to control for all elements of treatment except for the chemical ingredient of interest. Thus, the overall effects of treatment are the sum of the specific effect of the chemical agent plus the effects of 'extraneous' or non-specific factors such as patient expectancy, hope, and aspects of the doctor-patient relationship. With psychotherapeutic treatments, it is less clear which aspects of treatment are specific and which are non-specific. Furthermore, designing a credible control treatment that contains only the non-specific elements of the treatment package is inherently difficult.

(a) Types of control groups

A variety of types of control conditions have been implemented in psychotherapy efficacy studies. One common form of control group is the 'waiting-list' control. Patients are randomly assigned to either the experimental group or to delayed treatment. While such a control condition appears to control for the passage of time (i.e. patients generally improve over time even without treatment), assignment to a waiting-list control condition is likely to immediately change patient hope and expectations about relief of their problem. Patients in the waiting-list condition will not expect to improve until they actually receive treatment, while patients assigned to the experimental condition will likely be more hopeful about change. Thus, the treatment and control conditions are not balanced with regard to expectancy or other non-specific factors such as regular meetings with a competent caring professional. The potential superiority of the experimental condition over the control condition cannot be attributed to the hypothesized active ingredients in the experimental condition. Nevertheless, such waiting-list controls can serve as a useful initial step in evaluating a treatment.

A scientifically stronger type of control group for psychotherapy evaluation studies is one that involves regular meetings with a psychotherapist, but does not contain certain important hypothesized active elements. Some forms of supportive psychotherapy have been used as a control group in this regard. For example, in evaluating the efficacy of cognitive behavioural therapy for posttraumatic stress reactions, a supportive therapy control condition has been successfully used.[9] With some psychotherapy treatments, however, the supportive elements are an important part of the overall treatment model and are hypothesized to be curative in their own right. In this case, a supportive psychotherapy control condition is not appropriate for evaluating the efficacy of the overall package. In fact, sometimes supportive therapy has been found to be superior to other therapies.[10]

Comparative designs

An alternative to attempting to create an adequate psychological 'placebo' control is to only compare active treatments rather than comparing active treatments to 'placebo' control groups. However, rather than solve the problem of how to evaluate the efficacy of a psychotherapeutic intervention, comparative designs simply change the scientific question from 'Is this treatment efficacious?' to 'Which treatment works best?' Comparative designs are generally only informative if one active treatment proves to be more effective than another. If, as is commonly the case in psychotherapy studies, both active treatments produce equivalent results, the investigator is left not knowing whether both treatments are effective, or whether both treatments are non-effective (beyond the effects of non-specific elements).

Dismantling designs

Dismantling designs are an alternative to psychological placebos, waiting lists, or comparative designs. In a dismantling design, the full clinical treatment package is compared with the full package minus one element in order to establish which elements are necessary and sufficient for change. Variations on this theme include 'additive' or constructive designs that examine whether adding a new element enhances the efficacy of a treatment package.

As an example, dismantling designs have been usefully applied in the evaluation of the behavioural treatment of obsessive-compulsive disorder. The full treatment package involving exposure and response prevention techniques has been compared with each of the individual components of the package.[11] Patients were randomly assigned to either the full exposure and response prevention package, exposure alone, or response prevention alone. Those receiving the full package improved significantly more than did the patients in either of the control treatments. At follow-up assessments, 80 per cent of patients receiving the full package remained improved at follow-up, whereas only 27 per cent of those in the single components groups remained 'improved'.

The advantage of dismantling strategies in psychotherapy research is clear. Causal statements about differences in improvement between treatment conditions can be made, since all factors (including non-specific elements) except one are held constant, and the problems involved in other types of control groups are avoided. These designs, in general, should probably be used more often than they are. Some psychotherapies, however, may not easily lend themselves to such dismantling strategies. Moreover, it may be premature to attempt to dismantle a treatment package when questions about the efficacy of the whole package need to be resolved first.

Comparisons with medication

One type of design for the evaluation of psychotherapy that has increased in importance consists of comparisons to medication treatments. An example of this type of study is a recent investigation comparing cognitive therapy, medication (paroxetine), and pill placebo in the treatment of moderate to severe major depressive disorder.[12] The rationale for this design is that the medication group provides the standard reference condition with established efficacy. The pill-placebo group allows for establishing that the particular sample in the study showed the typical medication-placebo difference that would be expected (in other words, the sample was not unusual), and controls for some of the non-specific effects of the psychotherapy (for example, regular visits to a professional, positive expectancies for change). While the pill placebo is clearly not a perfect control condition for psychotherapy, it serves as a practical function—i.e. if a specific psychotherapy is not better than a pill placebo, should the psychotherapy be pursued as a treatment option?

Attribute by treatment interactions

An emerging emphasis in psychotherapy outcome research is the investigation of attribute by treatment interactions. Partly influenced by the common finding of no differences between psychotherapies, as well as the desire to make more specific clinical recommendations, investigators have hypothesized that certain matches of client characteristics and treatment modalities produce superior outcomes. One initial problem with pursuing potential matches of patient characteristics with the type of treatment is that there are a large number of potential variables (for instance, various combinations of diagnosis, therapist, treatment, patient's problem, setting).

The investigation of patient-treatment matches has intuitive clinical appeal, as most therapists believe that some patients seem to 'fit' a form of treatment better than others. Research on patient-treatment matching, however, is inherently difficult, particularly because large sample sizes are generally needed to adequately test interaction effects. Two of the largest randomized clinical trials ever performed with psychotherapeutic treatments have failed to find much support for specified patient-treatment interactions.[13,14] Thus, it remains to be seen whether aptitude by treatment-interaction designs will provide useful information about psychotherapy outcome.

Naturalistic studies

Naturalistic designs are used to examine issues such as the effectiveness of treatments in real-world setting with the types of patients that typically seek treatment. For example, the effectiveness of cognitive behavioural therapy for panic disorder and agoraphobia has been examined in a naturalistic study.[15] The authors report that outcomes were better for patients receiving naturalistic cognitive behaviour therapy, compared to a non-randomized wait-list control group. However, the overall outcomes for cognitive behavioural therapy were not as robust as previously seen in controlled clinical trials. While this form of naturalistic study has several advantages, including the fact that the data are drawn from actual clinical services in the real world, such data do not add to an understanding of which forms of psychotherapy work and which do not, and they do not provide strong causal statements. Hypotheses generated from naturalistic studies can inform the planning of experiments that can make stronger statements about causality. In addition, naturalistic studies provide useful descriptive information about the service delivery system.

Strategies for assessing psychotherapy outcome

Once a particular experimental design has been decided upon, the next crucial question in evaluating psychotherapy pertains to the selection of instruments for the study. Instruments include those needed to adequately select a patient population, such as measures of psychiatric diagnoses and initial symptomatology. Therapeutic change is often evaluated on a broad range of measures including dimensional measures of symptoms, personality, self-esteem, quality of life, and functioning in a variety of areas (for example, social and occupational functioning). In addition, the impact of treatment on specific theoretical constructs (mediators) can be examined. Discussion of each of these domains follows.

Patient selection

(a) Diagnosis

Because the identification of effective therapies for specific disorders is an important research priority from a public health point of view, an accurate and thorough assessment of these disorders is central to the evaluation process. The most widely used instrument for assessing the major DSM-IV Axis I diagnoses in research settings is the Structured Clinical Interview for DSM-IV.[16] The format of this semi-structured interview, with questions grouped by criteria and by diagnosis, allows an experienced clinical evaluator to assign diagnosis as the interview progresses. It is most commonly used to select subjects according to study inclusion or exclusion criteria or to characterize a study population. It can also be used to document change in diagnostic status post-treatment or over the course of a longitudinal study.

(b) Symptom measures

Whereas categorical measures are critical for the selection of a relatively homogeneous patient sample at intake, the evaluation of patient improvement requires the measurement of symptoms and functioning on a continuum. These continuous measures of the amount, timing, and nature of change are typically the primary outcomes of intervention studies. They include ratings of a single construct representing a core feature of the disorder, scales which cover the range of symptoms present in a general diagnostic category (e.g. rating scales of 'depression' or 'anxiety'), and measures that cut across many diagnoses and are indicative of overall psychopathology or symptomatology. At each of these levels of assessment, one can find both clinician-rated and self-report tools; for practical reasons, the self-report method predominates.

Measures of therapeutic change

(a) Dimensional assessment of core symptoms

In general, the more circumscribed and behavioural the problem, the simpler the assessment method. For some disorders, investigators have found that single-item measures of target symptoms suffice. For example, treatment studies of panic disorder and bulimia rely on self-reporting the number of 'episodes' that occur in a well-defined interval, such as the past week or the past month. Daily diaries are often used to facilitate recall and to enhance accuracy. Obsessive-compulsive symptoms can be assessed by Likert-type ratings of the severity of compulsions and obsessions, and the improvement of specific phobias by ratings of fear and avoidance. Severity ratings can be completed by both the patient and an independent evaluator.

For most disorders, core symptom measures of greater psychometric sophistication are needed to supplement simpler methods. The Penn State Worry Questionnaire,[17] for example, assesses the central feature of generalized anxiety disorder, and the Yale-Brown Obsessive-Compulsive Scale[18] measures the various symptoms that occur with obsessive compulsive disorder. In eating disorder research, the self-report EDI-2[19] and the Eating Disorder Examination[20] are commonly used.

In depression research, the Inventory of Depressive Symptomatology[21] is the result of efforts to develop a continuous measure of the nine DSM symptoms for major depressive episode. Other widely used scales include the interview-based Hamilton Rating Scale for Depression[22] and the self-report Beck Depression

Inventory,[23] although these instruments do not map directly on to DSM criteria. The severity of specific manic symptoms can be assessed with either the Bech-Rafaelsen Mania Scale[24] or the Young Mania Rating Scale.[25]

(b) General measures of anxiety, depression, and other symptoms

The Hamilton Rating Scale for Depression and the Beck Depression Inventory are often administered concurrently to provide clinician and patient perspectives on the level of depression, and are used both as primary outcome measures in studies of depression, and as secondary measures in treatment studies of other disorders. Because of the comorbidity of Axis I disorders in general, and the interrelatedness of anxiety and depression in particular, it is useful to include assessments of both depression and anxiety. General measures of anxiety include the Hamilton Anxiety Rating Scale[26] and the Beck Anxiety Inventory.[27] These general measures are used with a variety of diagnostic groups, but it is important to remember that they have varying relevance to different disorders within a given diagnostic group. For example, a score on the Beck Anxiety Inventory might be a good index of the severity of the generalized anxiety disorder, but be less informative about specific phobias or obsessive-compulsive behaviour.

(i) Psychotic symptoms

In psychological treatment studies of severe depression, mania, and psychotic disorders, it is often necessary to assess the level of psychotic symptomatology. The Brief Psychiatric Rating Scale[28] is recommended for this purpose. It has items covering five symptom clusters: thinking disturbance, anxious depression, withdrawal/retardation, hostile/suspiciousness, and agitation/excitement. As with other clinician-rated scales, it should be administered by an experienced clinician who has received standardized training on this instrument.

(ii) Substance use problems

Because of the frequent comorbidity of substance use problems with other Axis I disorders, it is important to evaluate the level of drug and alcohol use when screening patients for study enrollment or when characterizing a sample. The substance use modules of the DSM-IV Structured Clinical Interview are frequently used for this purpose. The Addiction Severity Index[29] has the advantage of providing more comprehensive and detailed information on problem areas associated with abuse, and yields scores that can be compared across patients, and from the beginning to the end of treatment. Because it is a rather time-consuming interview, it is recommended only in instances when the rates of substance use and related problems are expected to be high and their measurement is a research priority.

(c) Measure of global psychological functioning/ psychopathology

The rationale for using measures that cover a broad range of psychopathology is that they characterize the sample in terms of associated symptoms (i.e. symptoms other than those of primary interest) and provide a global measure of subjective distress. They have also proved to be quite useful in detecting treatment-related changes in evaluations of diverse psychotherapies. A popular self-report measure of general psychopathology is the SCL-90-R (and its abbreviated version, the Brief Symptom Inventory).[30] The SCL-90-R is a 90-item scale; the Brief Symptom Inventory is composed of 53 items. Both yield nine symptom dimensions and three global indices of distress. Because correlations between the two instruments are very high, it is recommended that the latter be used when time is an issue or when this measure is included as part of a larger core battery of assessment.

(d) Measures of self

Many psychotherapies explicitly attempt to improve self-esteem, self-concept, and self-confidence, and therefore it is relevant to examine the extent to which treatment successfully impacts on these domains. One of the oldest and most widely used measures of self-esteem is the Rosenberg Self-Esteem Scale,[31] an easily administered 10-item Likert-type scale yielding a uni-dimensional indicator of global self-esteem. More recent work in this area aims to distinguish other self-related constructs (such as self-concept) from self-esteem, develop more theoretically based multifactorial models, and improve upon the psychometrics of earlier measures. Some resulting scales, such as the Beck Self-Concept Test[32] or the Selves Questionnaire,[33] are appropriate for patients with a fairly wide range of psychiatric diagnoses.

(e) Personality assessment

Personality variables typically appear in evaluations of psychological treatment as either primary outcomes (as in the study of psychotherapy for personality pathology), or as prognostic indicators in studies of other Axis I disorders. They might also be included as part of a larger effort to thoroughly describe the patient sample. There are two ways to approach the evaluation of personality within these contexts: determination of the presence or absence of a DSM personality disorder, and dimensional ratings of personality features.

(i) Categorical: DSM-IV Axis II

Treatments that target specific personality disorders tend to rely on the first approach to assessment, namely the use of interviewer-administered instruments which assess criteria for the 10 specific personality disorders listed in the DSM-IV. The Structured Clinical Interview for DSM-IV Axis II[34] was designed for this purpose. A positive feature of this assessment is its efficiency in eliciting the information required to assign Axis II diagnoses, especially when used in conjunction with the self-report Personality Disorder Questionnaire.

(ii) Dimensional

Dimensional methods of assessing personality arise from either theoretical or factor-analytical models of the essential elements of personality structure. The Five Factor Model[35] proposes that personality, in both patient and non-patient samples, can be measured along five dimensions: neuroticism, extraversion, openness to experience, agreeable, and conscientiousness. Extremes of these traits define personality pathology. This model and several instruments used to generate scores on these factors, including the self-report NEO Personality Inventory and NEO-PI-R[36] and a semistructured interview (the Structured Interview of the Five-Factor Model of Personality[37]), have received considerable empirical support in both clinical and non-clinical samples.

Another dimensional measure of personality is the SWAP-200.[38] This instrument consists of a set of 200 personality-descriptive statements designed to be used by a clinician who knows a given patient well. The clinician arranges the 200 statements into eight categories, from those that are not descriptive to those that are highly descriptive of the patient. These scores are then used to generate both individualized, ideographic case descriptions as well as dimensional scores for the 10 personality disorders included in DSM–IV.

The categorical and dimensional methods are not entirely mutually exclusive and both are valuable in ongoing research into the phenomenology, aetiology, correlates, and treatment of personality disorder.

(f) Measures of functioning and quality of life

In recent years, the definition of mental health has been broadened to encompass its more positive, and somewhat paradoxically, its more negative aspects. Mental health is now viewed as more than the absence of illness, but the illness itself is increasingly seen as chronic and quite disabling. This has prompted investigators to turn their attention to outcomes other than symptom change—to evaluate fully the effectiveness of a psychotherapeutic intervention means to document its broad effect on a number of relevant domains including social/interpersonal functioning, work functioning, and quality of life.

(i) Overall functioning

Incorporated into the DSM system as a separate axis (Axis V), the Global Assessment of Functioning scale[39] (GAF) is the most widely used measure of psychosocial impairment. Clinicians rate on a scale of 1 to 100 the current level of functioning, or if desired, the highest or lowest level of functioning within some designated time period. General well being and functioning are also assessed with scales that have been developed for the routine outcome assessment in clinical practice, rather than for efficacy trials. An example is the OQ-45,[40] a self-report measure consisting of three subscales representing broad content areas: (1) symptom distress, (2) interpersonal relations, and (3) social role (dissatisfaction and distress in tasks related to work, family roles, and leisure life).

(ii) Social and occupational functioning

The assessment of role performance has a fairly long history in psychotherapy research. The Social Adjustment Scale (SAS)[41] was developed to document the level in functioning in six areas: work as worker, 'housewife', or student; social activities; relationship with family; relationship with spouse; parental responsibilities; member of a family unit. There is much documentation of its psychometric properties and clinical utility, and it has been used with a wide range of adult outpatients. A unique feature of the Longitudinal Interval Follow-up Evaluation (LIFE)[42] is that it was designed to collect information on psychosocial functioning (as well as on diagnosis and symptoms) over longer periods of time. In the hands of trained evaluators, it is a structured interview that has been shown to have good reliability.

The concern among both scientists and health-care managers about the cost-effectiveness of specific treatments has generated interest in more comprehensive assessments of work functioning and productivity. Endicott and Nee developed the Endicott Work Productivity Scale,[43] a self-report instrument containing 25 items designed to be sensitive to more subtle differences among patients in work attitudes and behaviour.

Two self-report instruments measuring interpersonal difficulties have been used in studies of psychotherapy: The Inventory of Interpersonal Problems[44] and the Dyadic Adjustment Scale.[45] The Inventory of Interpersonal Problems is a 127-item self-report measure with high internal consistency and test-retest reliability of each of the six subscales: assertive, social, intimate, submissive, responsible, and controlling. The Dyadic Adjustment Scale is a 32-item scale designed to assess the severity of relationship discord in married and unmarried cohabiting couples, with higher scores indicating better adjustment. Responses on the Dyadic Adjustment Scale discriminate between distressed and non-distressed couples and yield a total score based on the factors of dyadic consensus, dyadic cohesion, dyadic satisfaction, and affectional expression.

(iii) Quality of life

The most promising of the quality-of-life measures for mental health outcome research are those based on a broad definition of quality of life—covering role functioning and social-material conditions as well as life satisfaction or well being—and which can be applied across disorders and across treatments. However, more widespread use and testing of quality-of-life instruments is needed to establish their utility and to help resolve a number of important measurement issues, including the value of more 'objective' indices of quality of life and the relationship between symptoms and quality-of-life judgments.

(g) Utilization of treatment services

The impact of a treatment on usage of other health-related services is a factor to be considered in determining the cost-effectiveness of that treatment. One broad instrument designed to assess changes in service usage over the course of treatment is the Treatment Services Review.[46] This is a 5-minute interview that documents the number and types of treatment services received during rehabilitation from substance abuse, but it also can be adapted for use with other clinical populations.

(h) Theory-based measures

Evaluation of the hypothesized important psychological constructs of a particular psychotherapy can serve as outcome measures in their own right or as mediators of change in symptoms and functioning. For example, the cognitive model of depression holds that distorted cognitions about the self and world are responsible for generating and maintaining negative emotions. Measures of depressogenic cognitions are therefore included as outcomes and mediators of symptom change in studies of cognitive therapy for depression. The Hopelessness Scale[47] is a 20-item self-report scale that assesses the hopelessness and pessimism associated with suicidal ideation and intent. The Dysfunctional Attitudes Scale[48] is a 40-item index of general attitudes and beliefs hypothesized by Beck and colleagues to underlie a propensity for depressive thinking, whereas the Automatic Thoughts Questionnaire[49] covers 30 negative thoughts proposed to occur during a symptomatic depressed state.

In regard to psychodynamic psychotherapy, theory-specific mediators include measures of core conflicts[50] and self-understanding.[51]

Conclusions

Of all the treatments in medicine and psychiatry in particular, the evaluation of psychotherapy offers some of the greatest challenges to researchers. A wide variety of research designs and instruments have been employed, with distinct advantages and disadvantages to each. No one study of psychotherapy can answer all the important questions that need to be asked. Given the complexity of research on psychotherapy, knowledge accumulates slowly. Despite the problems, complexities, and slow pace of scientific advance,

psychotherapy outcome research now has the tools and emerging findings to begin to influence the practice of psychotherapy and mental health treatments in general. We expect that in the next decade methodological advances in the evaluation of psychotherapy will lead to a stronger link between research and practice.

Further informaion

Lambert, M. (2004). *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th edn). John Wiley and Sons, New York.

Society for Psychotherapy Research. http://www.psychotherapyresearch.org

Rush, A.J. (2007). *Handbook of psychiatric measures* (2nd edn). American Psychiatric Publishing, Washington, DC.

References

1. Eysenck, H.J. (1952). The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology*, **16**, 319–24.
2. Smith, M.L., Glass, G.V., and Miller, T.I. (1980). *The benefits of psychotherapy*. Johns Hopkins University Press, Baltimore, MD.
3. Castonguay, L.G., and Beutler, L.E. (2006). (eds). *Principles of therapeutic change that work*. Oxford University Press, New York.
4. Nathan, P., and Gorman, J. (in press) (eds.). *Treatments that Work,* Oxford Unversity Press (3rd edn.), New York.
5. Kazdin, A. (1994). Methodology, design, and evaluation in psychotherapy research. In *Handbook of psychotherapy and behavior change* (ed. A.E. Bergin and S.L. Garfield), pp. 19–71. Wiley, New York.
6. Crits-Christoph, P. and Gallop, R. (2006). Therapist effects in the TDCRP and other psychotherapy studies. *Psychotherapy Research,* **16**, 178–81.
7. Crits–Christoph, P., Baramackie, K., Kurcias, J., *et al*. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research*, **1**, 81–91.
8. Kazdin, A. (1982). *Single–case research designs: methods for clinical and applied settings*. Oxford University Press, New York.
9. Blanchard, E.B., Hickling, E.J., Devineni, T., *et al*. (2003). A controlled evaluation of cognitive behavioral therapy for posttraumatic stress in motor vehicle accident survivors. *Behaviour Research and Therapy* **41**, 79–96.
10. McIntosh, V.V., Jordan, J., Carter, FA., *et al*. (2005). Three psychotherapies for anorexia nervosa: a randomized, controlled trial. *American Journal of Psychiatry,* **162**, 741–7.
11. Foa, E., Steketee, G., Grayson, J., *et al.* (1984). Deliberate exposure and blocking of obsessive–compulsive rituals: immediate and long-term effects. *Behavior Therapy*, **15**, 450–72.
12. DeRubeis, R.J., Hollon, S.D., Amsterdam, J.D., *et al. (2005)*. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry,* **62**, 409–16.
13. Project MATCH Research Group (1997). Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol*, **58**, 7–29.
14. Crits–Christoph, P., Sigueland, L., Blaine, J., *et al.* (1999). Psychosocial treatments for cocaine dependence: National Institute on Drug Abuse Collaborative Cocaine Treatment Study. *Archives of General Psychiatry*, **56**, 493–502.
15. Rosenberg, N.K., and Hougaard E. (2005). Cognitive behavioural group treatment of panic disorder and agoraphobia in a psychiatric setting: A naturalistic study of effectiveness. *Nordic Journal of Psychiatry.* **59**, 198–204.
16. First, M.B., Spitzer, R.L., Gibbon, M., *et al.* (1994). *Structured Clinical Interview for Axis–I DSM–IV Disorders. Patient edition (SCID-II, version 2.0)*. Biometrics Research Department, New York State Psychiatric Institute.
17. Meyer, T.J., Miller, M.L., Metzger, R.L., *et al.* (1990). Development and validation of the Penn State Worry Questionnaire. *Behavior Research and Therapy*, **28**, 487–95.
18. Goodman, W.M., Price, L.H., Rasmussen, C.A., *et al.* (1989). The Yale–Brown Obsessive–Compulsive Scale. I: development, use, and reliability. *Archives of General Psychiatry*, **46**, 1006–11.
19. Garner, D.M. (1991). *Eating Disorder Inventory–2 manual.* Psychological Assessment Resources, Odessa, FL.
20. Fairburn, C. G., and Cooper, Z. (1993).The Eating Disorder Examination (12th edn.). In *Binge eating: Nature, assessment, and treatment* (eds. C.G.Fairburn and G.T. Wilson) pp. 317–60. Guilford Press, New York.
21. Rush, A.J., Carmody, T.J., Ibrahim, H.M., *et al.* (2006). Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. *Psychiatric Services*, **57**, 829–37.
22. Hamilton, M. (1960). A rating scale for depression. *Journal of Neurological and Neurosurgical Psychiatry*, **23**, 56–62.
23. Beck, A.T., Steer, R.A., and Brown, G.K. (1996). *BDI-II manual* (2nd edn). Psychological Corporation, San Antonio, TX.
24. Bech P. (2002). The Bech-Rafaelsen Mania Scale in clinical trials of therapies for bipolar disorder: a 20-year review of its use as an outcome measure. *CNS Drugs*, **16**, 47–63.
25. Young, R.C., Biggs, J.T., Ziegler, V.E., *et al.* (1978). A rating scale for mania: reliability, validity and sensitivity. *British Journal of Psychiatry*, **133**, 429–35.
26. Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology*, **32**, 50–5.
27. Steer, R.A., Beck, A.T. (1997). In *Evaluating stress: A book of resources* (eds. C.P. Zalaquett and R.J Wood), pp. 23–40. Scarecrow Education, Lanham, MD.
28. Thomas, A., Donnell, A.J., and Young, T.R. (2004). Factor structure and differential validity of the expanded Brief Psychiatric Rating Scale. *Assessment*, **11**, 177–87.
29. McLellan, A.T., Kushner, H., Metzger, D. *et al.* (1992). The fifth edition of the addiction severity index: historical critique and normative data, *Journal of Substance Abuse Treament*, **9,** 199–213.
30. Derogatis, L.R., and Fitzpatrick, M. (2004). The SCL-90-R, the Brief Symptom Inventory (BSI), and the BSI-18 In *The use of psychological testing for treatment planning and outcomes assessment: Volume 3: Instruments for adults* (3rd edn.) (ed. M.E. Maruish). (pp. 1–41). Lawrence Erlbaum Associates Publishers, Mahwah, NJ.
31. Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.
32. Beck, A.T., Steer, R.A., Brown, G., *et al.* (1990). The Beck Self-Concept Test. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, **2**, 191–7.
33. Higgins, E.T., Bond, R.N., Klein, R., *et al.* (1986). Self-discrepancies and emotional vulnerability: how magnitude, accessibility, and type of discrepancy influence affect. *Journal of Personality and Social Psychology*, **51**, 5–15.
34. First, M.B., Spitzer, R.L., Gibbon, M., *et al.* (1994). *Structured Clinical Interview for DSM-IV Axis-II Personality Disorders (SCID-II, version 2.0)*. Biometrics Research Department, New York State Psychiatric Institute.
35. McCrae, R.R., and Costa, P.T. (2003). *Personality in adulthood: A five-factor theory perspective* (2nd edn.) Guilford Press, New York, NY.
36. McCrae, R.R., and Costa, P. T. (2005). The NEO-PI-3: A More Readable Revised NEO Personality Inventory. *Journal of Personality Assessment*, **84**, 261–70.
37. Trull, T.J., and Widiger, T.A. (2002). The structured interview for the five factor model of personality (SIFFM) In *Big five assessment* (eds. B. de Raad & M. Perugini). (pp. 148–66). Ashland, OH: Hogrefe & Huber Publishers.
38. Westen, D., and Muderrisoglu, S. (2006). Clinical assessment of pathological personality traits. *American Journal of Psychiatry*, **163**, 1285–7.

39. American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
40. Lambert., M.J., Gregersen A.T., and Burlingame, G.M. (2004). The Outcome Questionnaire-45. In *The use of psychological testing for treatment planning and outcomes assessment: Volume 3: Instruments for adults* (3rd edn) (ed. M.E. Maruish), pp. 191–234. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
41. Weissman, M.M., Klerman, G.L., Paykel, E.S., *et al.* (1974). Treatment effects in the social adjustment of depressed patients. *Archives of General Psychiatry*, **30**, 771–8.
42. Keller, M.B., Lavori, P.W., Friedman, B., *et al.* (1987). The longitudinal interval follow-up evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry*, **44**, 540–8.
43. Endicott, J. and Nee, J. (1997). Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. *Psychopharmacology Bulletin*, **33**, 13–16.
44. Horowitz, L.M., Rosenberg, S.E., Baer, B.A, *et al.* (1988). Inventory of Interpersonal Problems: psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, **56**, 885–92.
45. Spanier, G.B. (1976). Measuring dyadic adjustment: new scales for assessing the quality of marriage and similar dyads. *Journal of Marital and Family Therapy*, **38**, 15–38.
46. McLellan, A.T., Alterman, A.I., Cacciola, J., *et al.* (1992). A new measure of substance abuse treatment. Initial studies of the treatment services review. *Journal of Nervous and Mental Disease*, **180**, 101–10.
47. Beck, A.T. and Steer, R.A. (1988). *Manual for Beck Hopelessness Scale*. Psychological Corporation, San Antonio, TX.
48. Weissman, A.N. (1979).The dysfunctional attitudes scale: a validation study. *Dissertation Abstracts International*, **40**, 1389–90.
49. Hollon, S.D. and Kendall, P.C. (1980). Cognitive self-statements in depression: development of an automatic thoughts questionnaire. *Cognitive Therapy and Research*, **4**, 88–100.
50. Luborsky, L. and Crits–Christoph, P. (1998). *Understanding transference: the core conflictual relationship theme method*. American Psychological Association, Washington, DC.
51. Connolly, M.B., Crits–Christoph, P., Shelton, R.C., *et al.* (1999). The reliability and validity of a measure of self–understanding of interpersonal patterns. *Journal of Counseling Psychology*, **46,** 472–82.

6.2

Somatic treatments

Contents

6.2.1 General principles of drug therapy in psychiatry

J. K. Aronson

The successful use of psychotropic drugs demands an understanding of their pharmaceutical, pharmacokinetic, and pharmacodynamic properties.

- *Pharmaceutical properties*: Pharmaceutical formulations can be manipulated to produce different durations of action, for example the use of oily emulsions of antipsychotic drugs in depot formulations.
- *Pharmacokinetic properties*: Pharmacokinetics is the mathematical description of the disposition of drugs in the body by absorption, distribution (to plasma proteins and tissues), and elimination (usually by hepatic metabolism and renal excretion). Differences in drug disposition determine differences in dosage regimens and are important for drug interactions.
- *Pharmacodynamic properties*: Pharmacodynamics is the study of the pharmacological actions of drugs and how actions at the molecular level are translated, via actions at cellular, tissue, and organ levels, into therapeutic or adverse effects. The known pharmacological actions of psychotropic drugs are not necessarily the actions that produce their therapeutic or adverse effects.

Dosage regimens

A drug dosage regimen is a recipe for drug administration, intended to produce the desired therapeutic effect with a minimum of unwanted effects. It is described in terms of the pharmaceutical formulation, the dose, and the frequency and route of administration used. The duration of administration is also important.

In treating any condition, it is best to learn initially to use a few drugs, preferably well-established ones, and to expand one's repertoire with increasing experience.

The choice of drug depends firstly on the indication—obviously an antidepressant will be the drug of choice for a patient with depression, if drug therapy is thought to be required. The choice of

antidepressant will depend on features of the disease and other factors. For example, some antidepressants are more sedative and anxiolytic than others, and can be helpful in patients who are agitated. The avoidance of adverse effects or interactions can also dictate the choice; for example, tricyclic antidepressants should be avoided in men with prostatic hyperplasia and selective serotonin reuptake inhibitors (SSRIs) should not be used in children, because of the increased risk of suicidal ideation.

It is usual to start therapy with published dosage recommendations, generally beginning at the lower end of the recommended dosage range and monitoring for a therapeutic effect. A common error is to give a starting dose of a drug and then to add or substitute another drug if the first does not work. This is usually bad practice. If the desired effect does not occur with the initial dosage, increase the dose gradually until the effect occurs or the upper limit of the recommended range is reached (although adverse effects may limit this process). Only then should another drug be tried. Sometimes a poor response is due to poor adherence to therapy; careful explanation of the condition and the need for therapy helps.

Psychotropic drugs can be given orally or parenterally, and as immediate-release or modified-release formulations. Most drug administration is oral, but parenteral therapy can be useful to guarantee administration (e.g. depot formulations in schizophrenia) and for a quicker onset of action (e.g. in the treatment of acute mania). Modified-release products are used for long-term therapy. They can be given less often and produce a smoother profile of blood concentrations (Fig. 6.2.1.1).[1] The advantage of intramuscular modified-release (depot) formulations of antipsychotic drugs is that drug delivery can be ensured by supervised infrequent administration (say every 2 weeks). Different modified-release formulations of the same compounds have different release characteristics and are not interchangeable; for example, when prescribing a modified-release oral formulation of lithium always give the patient the same formulation and specify the brand name on the prescription.

Combination formulations (e.g. a phenothiazine plus a tricyclic antidepressant in a single tablet) do not allow flexibility of prescribing and should generally be avoided. Important exceptions include combination analgesic formulations (e.g. co-codamol, which contains paracetamol plus codeine) and combinations of levodopa with a dopa decarboxylase inhibitor (benserazide or carbidopa).

Treatment of children

There are no uniform rules for determining dosage regimens in children. Pharmacokinetics and pharmacodynamics are different for some drugs but not others.

- Absorption is not greatly different from absorption in adults.
- The distribution of water-soluble drugs is different, but psychotropic drugs are lipid-soluble.
- Protein binding is reduced in neonates; phenytoin is affected.
- Hepatic oxidative metabolism and glucuronide conjugation are deficient in neonates, and mature at variable rates; this is important for psychotropic drugs.
- Glomerular and renal tubular functions are immature in neonates and take about 6 months to reach adult values.

If a child needs a psychotropic drug, consult the manufacturer's literature and always start with a low dosage.

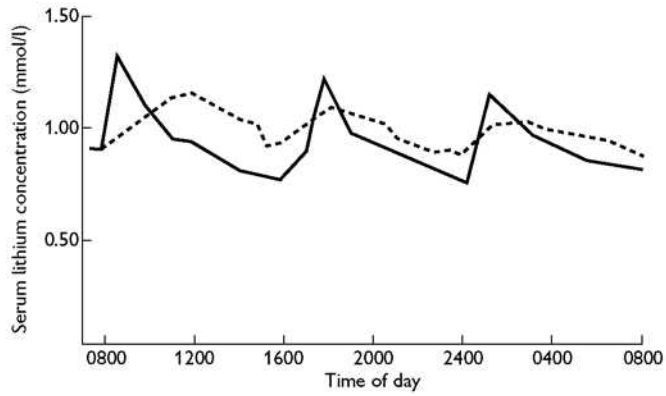


Fig. 6.2.1.1 Administration of lithium in immediate-release and modified-release oral formulations. The immediate-release formulation (solid line) produces rapid peaks of serum concentration and large fluctuations during a dosage interval. In contrast, the modified-release formulation (dotted line) is more slowly absorbed but produces much less fluctuation in serum concentrations. Note also that the apparent half-life of lithium is longer after administration of the modified-release formulation; this is not the true half-life of lithium, but the half-life of its release from the modified-release formulation. (Adapted from A. Amdisen, Variation of serum lithium concentration during the day in relation to treatment control, absorptive side effects and the use of slow-release tablets, *Acta Psychiatrica Scandinavica*, **207**, 55–7, copyright 1969, John Wiley & Sons, Inc.)

Treatment of elderly people

Pharmacokinetic differences in old age are more predictable than in children, but pharmacodynamic changes are variable.

- Absorption is not greatly affected.
- Elderly people have less body fat, and so lipid-soluble drugs may be more highly concentrated in the brain; however, this effect varies unpredictably from drug to drug (e.g. the apparent volume of distribution of diazepam is increased while that of nitrazepam is not).
- Protein binding is reduced in elderly people; phenytoin is affected.
- Hepatic metabolism is reduced in frail but not in fit old people; this effect is proportional to liver size.
- Renal function is impaired with age; use creatinine clearance, measured or estimated (not eGFR), as a guide.
- Inappropriate polypharmacy is common in old people, increasing the risk of drug interactions.

When treating an elderly person with a psychotropic drug always start with a low dosage and increase dosages more slowly.

Pregnancy and breast feeding

Anticonvulsants are teratogenic.[2] For example, sodium valproate has been associated with spina bifida, cardiac malformations, hypospadias, anomalies of the brain and face, coarctation of the aorta, and limb reduction defects.[3]

Few other psychoactive drugs are teratogenic. However, most of them cross the placenta and some can cause withdrawal symptoms in the neonate. The teratogenicity of lithium has been overstated in the past; the main risk is cardiovascular teratogenicity, but although the risk of Ebstein's anomaly is increased, the absolute risk (0.05–0.1 per cent) is still very small;[4] nevertheless, some

advise that it should be avoided or used with caution in the first trimester of pregnancy,[5] and fetal sonography is recommended at 18–20 weeks after first-trimester exposure.[3]

Although most psychoactive drugs are lipid-soluble and therefore enter the breast milk, few do so in high enough amounts to trouble the neonate; if a neonate becomes drowsy while breast feeding, reduce the mother's dosage or stop breast feeding. Lithium appears in the breast milk and can be found in the serum of breast-fed babies in variable concentrations, up to half of those in the mother. Because neonates have immature renal function, some recommend avoiding breast feeding.[6] However, others consider that the benefits of breast feeding to mother and child outweigh the small risk of lithium toxicity.[4] The following advice has been given:[3]

- Educate the mother about the manifestations of toxicity.
- Explain the risks of dehydration.
- Consider partial or total formula supplements during episodes of illness or dehydration.
- Suspend breast feeding if toxicity is suspected.
- Check infant and maternal serum concentrations.

Pharmacokinetics—drug disposition

Most psychotropic drugs are rapidly and well absorbed after oral administration. However, drugs can be removed by various processes before they reach the systemic circulation. The fraction of drug that reaches the systemic circulation is called its systemic availability (or, more commonly, bioavailability).

After oral administration a formulation will generally disintegrate in the stomach and the drug it contains will dissolve in gastric contents. However, drugs are not generally absorbed in the stomach. After gastric emptying they are for the most part absorbed in the jejunum and ileum, and some are absorbed from the colon as well. During transit across the gut wall they may be metabolized by an oxidative isozyme of cytochrome P450, CYP3A4, and can be secreted back into the gut lumen by P glycoprotein. When they enter the portal circulation they may be eliminated by the liver. If hepatic metabolism is extensive, a large amount of drug will be removed during this first passage through the liver. For example, clomethiazole has extensive first-pass metabolism in the liver and its systemic availability is low (about 40 per cent); thus, intravenous doses are considerably lower than oral doses. In severe liver disease, such as cirrhosis, or when there is arteriovenous shunting, this presystemic metabolism is reduced and the systemic availability increases up to 90 per cent; oral doses of clomethiazole should be reduced in liver disease.[7]

In the systemic circulation drugs are bound to plasma proteins and distributed to the tissues. Protein binding is important for drugs that are highly bound (over 90 per cent) and not widely distributed to the body tissues; in those cases protein-binding displacement can result in a large rise in the amount of unbound drug available to the target tissue. This is important for phenytoin, which is 90 per cent bound to plasma albumin and has a low volume of distribution. The binding of phenytoin is reduced when the serum albumin concentration falls (in chronic liver disease, the nephrotic syndrome, protein malnutrition, or the third trimester of pregnancy), when binding to the protein is abnormal (in chronic renal insufficiency), or when another drug (e.g. sodium valproate) causes displacement. Acute displacement causes phenytoin toxicity, but only transiently, because in the case of phenytoin an increase in unbound concentration causes it to be more rapidly eliminated. When measuring plasma phenytoin concentrations in patients in whom protein binding is reduced, the target concentration (and the laboratory will measure total drug, i.e. bound plus unbound) is reduced (see Fig. 6.2.1.2).

In chronic renal insufficiency the protein binding of phenytoin is reduced. This leads to an increase in the unbound plasma (or serum) concentration relative to the total concentration; the target total concentration therefore falls. The shaded area shows the range of plasma phenytoin concentrations that one would generally aim to achieve (the target concentration range) in treating a patient with epilepsy. As renal function deteriorates (indicated here by an increase in serum creatinine concentration), the target range for plasma phenytoin concentration falls from 40–80 μmol/l when renal function is normal to 10–30 μmol/l in severe renal insufficiency.

After absorption and distribution most psychoactive drugs are cleared from the body by hepatic metabolism;[8] impaired liver function, if severe (for the liver has a large capacity), reduces their elimination, and dosages should be reduced.

Lithium is cleared solely by renal elimination and therefore the dosage should be reduced in proportion to the creatinine clearance. Since renal function falls with age, lithium dosages should be lower in older people.[9]

The half-life of a drug is a function of its clearance and its distribution volume: the slower the rate of clearance or the more extensive the distribution the longer the half-life. If a drug is given in a regular maintenance dose, the amount of drug in the body will gradually accumulate; however, as the amount in the body increases, the rate at which it is eliminated also rises, and eventually a plateau (or steady state) is reached when the amount eliminated during a dosage interval equals the dose (the maintenance dose). The time it takes to reach this steady state depends on the half-life

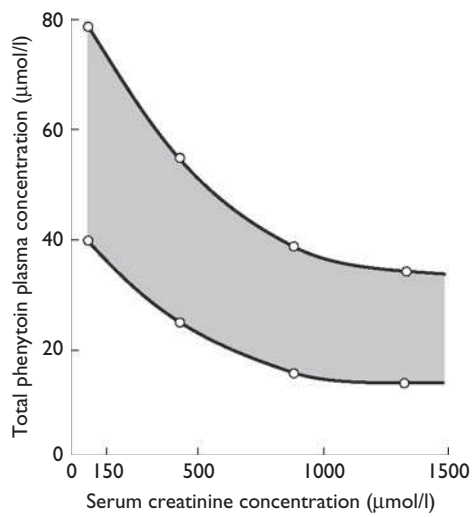


Fig. 6.2.1.2 In chronic renal insufficiency the protein binding of phenytoin is reduced. This leads to an increase in the unbound plasma (or serum) concentration relative to the total concentration; the target total concentration therefore falls. The shaded area shows the range of plasma phenytoin concentrations that one would generally aim to achieve (the target concentration range) in treating a patient with epilepsy. As renal function deteriorates (indicated here by an increase in serum creatinine concentration) the target range for plasma phenytoin concentration falls, from 40–80 μmol/l when renal function is normal to 10–30 μmol/l in severe renal insufficiency (I. Odar-Cederlöf, unpublished data.)

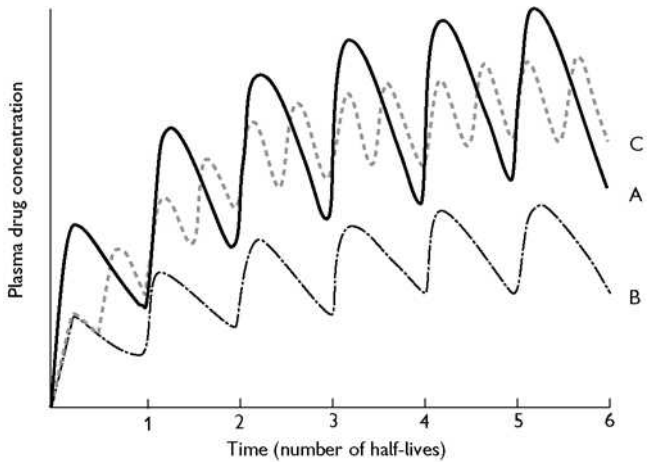


Fig. 6.2.1.3 Curve A—during the regular administration of a maintenance dose of a drug the amount of drug in the body rises after a dose, reaches a peak, and then falls as the drug is distributed to the tissues and eliminated. If another dose is given soon after the first, the plasma concentration will rise by the same amount as before but will fall faster after peaking, since most drugs obey first-order kinetics and the plasma concentration falls exponentially. Thus, when a drug is given repeatedly the mean plasma concentration rises more slowly with each successive dose, until eventually a steady state is reached, when the amount eliminated in a dosage interval is equal to the dose itself. This takes about four half-lives of the drug. Curve B represents the concentrations during administration of half the dose given at the same frequency. The time taken to reach steady state is the same in both cases, but the eventual steady-state concentration in case B is half that in case A, being proportional to the dose. Curve C represents the concentrations during administration of half the dose given twice as often (i.e. the total dose is unchanged). Neither the time taken to reach steady state nor the eventual mean steady-state concentration is affected. However, the fluctuations in plasma concentration during a dosage interval are reduced (cf. Fig. 6.2.1.1). (Adapted from Amdisen, A. Variation of serum lithium concentration during the day in relation to treatment control, absorptive side effects and the use of slow-release tablets, *Acta Psychiatrica Scandinavica*, **207**, 55–57. Copyright 1969, John Wiley & Sons, Inc.)

of the drug; about 94 per cent of the steady-state value will be reached after four half-lives (Fig. 6.2.1.3, curve A). For example, lithium has a half-life of about 24 h; after 4 days of maintenance therapy with the same regular dose a steady state will be reached; this does not depend on the dose or frequency of administration (Fig. 6.2.1.3, curves B and C). If a modified-release formulation is used and the half-life of absorption of the drug from the formulation is longer than the drug's own half-life, the longer (apparent) half-life will determine the time to steady state; for example, the apparent half-life of flupentixol after the administration of flupentixol decanoate is 17 days, compared with 36 h for flupentixol after oral administration. When using depot antipsychotic drugs, which have long half-lives of absorption, steady-state therapy should first be established with an ordinary formulation.

Curve A shows that during the regular administration of a maintenance dose of a drug the amount of drug in the body rises after a dose, reaches a peak, and then falls as the drug is distributed to the tissues and eliminated. If another dose is given soon after the first, the plasma concentration will rise by the same amount as before but will fall faster after peaking, since most drugs obey first-order kinetics and the plasma concentration falls exponentially. Thus, when a drug is given repeatedly the mean plasma concentration rises more slowly with each successive dose, until eventually a steady state is reached, when the amount eliminated in a dosage interval is equal to the dose itself. This takes about four half-lives of the drug. Curve B represents the concentrations during administration of half the dose given at the same frequency. The time taken to reach steady state is the same in both cases, but the eventual steady-state concentration in case B is half that in case A, being proportional to the dose. Curve C represents the concentrations during administration of half the dose given twice as often (i.e. the total dose is unchanged). Neither the time taken to reach steady state nor the eventual mean steady-state concentration is affected. However, the fluctuations in plasma concentration during a dosage interval are reduced (cf. Fig. 6.2.1.1). Kinetic characteristics of some psychotropic drugs are shown in Table 6.2.1.1.

Pharmacological actions of drugs

Psychotropic drugs interfere with neurotransmitter functions in several ways—via actions on neurotransmitter receptors, storage, release, reuptake, and metabolism. Transmembrane neurotransmitter receptors are broadly speaking of two types—ionotropic and metabotropic receptors. Ionotropic receptors (e.g. nicotinic acetylcholine, glycine, GABA, and NMDA, AMPA, and kainate receptors) incorporate ion channels in their structures and mediate rapid responses. Metabotropic receptors (e.g. G protein-coupled receptors such as adrenaline, noradrenaline, cannabinoid, dopamine, opioid, and serotonin receptors other than $5HT_3$ receptors) produce their effects via signal transduction systems, which activate second messengers or ion channels, and produce longer lasting responses.

Agonist action at a receptor

Agonists are substances that act by stimulating the action of a receptor.

Benzodiazepines bind to benzodiazepine receptors in the spinal cord, brainstem, cerebellum, limbic system, and cerebral cortex. These receptors are associated with receptors for the inhibitory neurotransmitter γ-aminobutyric acid (GABA), linked to a chloride channel.[10] The benzodiazepines enhance the action of GABA through its chloride channel, the presumed mechanism whereby they are anxiolytic and hypnotic. Some other hypnotics that have non-benzodiazepine structures also act via benzodiazepine receptors: zopiclone binds to the GABA–benzodiazepine receptor complex, but at a site different from that of benzodiazepines;[11] clomethiazole binds to a binding site distinct from those of benzodiazepines and barbiturates;[12] zolpidem binds to a subtype of binding site called BZ_1, found on GABA neurones in the sensorimotor cortex and extrapyramidal tracts.[10]

The triptans (such as sumatriptan, naratriptan, zolmitriptan), which are used to treat migraine, are agonists at 5-hydroxytryptamine ($5\text{-}HT_{1B/D}$) receptors, causing vasoconstriction. They are therefore contraindicated in patients with cardiovascular disease and in those with hemiplegic or basilar migraine because of the fear of stroke.[13]

Antagonist action at a receptor

Antagonists are substances that have no actions of their own at receptors and act by preventing the action of an agonist, usually an endogenous one.

The antipsychotic (neuroleptic) drugs are all antagonists at receptors for the endogenous neurotransmitter dopamine; this is thought to be the basis of their antipsychotic actions in the mesolimbic system (via D_1 and D_4 receptors) and undoubtedly produces their

Table 6.2.1.1 Pharmacokinetic information about some psychotropic drugs

Drug	Systemic availability (%)	Half-life (h)	Route of hepatic elimination[a]
Benzodiazepines			
Alprazolam	90	12	CYP3A
Chlordiazepoxide[b]	95	15	
Clobazam[b]	90	20	
Diazepam[b]	100	30	CYP2E1/2C19
Flurazepam[b]	30	3	CYP2E1
Lorazepam	90	15	
Nitrazepam	75	24	
Oxazepam	100	9	
Temazepam	95	10	
Antidepressonts			
Lithium	100	24	(Renally excreted)
Tricyclics			
Amitriptyline	50	20	CYP1A2/2D6
Clomipramine	50	20	CYP1A2/2D6/2C19
Desipramine	40	24	CYP2D6
Imipramine[b]	50	14	CYP1A2/2D6
Nortriptyline	60	36	CYP2D6
Tetracyclics			
Maprotiline	95	40	
Mianserin	25	16	CYP2D6
Triazolopyridines			
Trazodone	100	10	
Selective serotonin reuptake inhibitors			
Fluoxetine	95	48	CYP2D6/3A4
Fluvoxamine	90	20	CYP1A2/2C19/3A4
Paroxetine	Variable	24	CYP2D6
Sertraline	Low	24	CYP2D6
Monoamine oxidase inhibitors			
Phenelzine	High	1	Polymorphically acetylated
Tranylcypromine	Moderate	2	
Moclobemide	40	2	CYP2C19
Antipsychotic drugs			
Phenothiazines			
Chlorpromazine[b]	10	12	
Thioridazine	60	10	CYP2D6
Trifluoperazine	Low	14	
Butyrophenones			
Haloperidol	60	20	CYP2D6
Droperidol	75	2	
Thioxanthenes			
Flupentixol	40	36	
Zuclopenthixol	50	20	CYP2D6
Others			
Clozapine	50	12	
Risperidone	75	3/20[c]	CYP2D6

[a] CYP refers to isozymes of the cytochrome P-450 family of enzymes.

[b] All these drugs are partly metabolized to active metabolites, Some of the metabolites have long half-lives (e.g. diazepam is metabolized to desmethyldiazepam). Some benzodiazepines (e.g. clorazepate and prazepam) are completely metabolized to active metabolites with long half-lives.

[c] Extensive and poor metabolizers respectively.

adverse effects in the extrapyramidal tracts (via D_2 receptors). The so-called atypical antipsychotic drugs (including clozapine and risperidone) have little effect on D_2 receptors and less commonly cause extrapyramidal adverse effects.[14]

Flumazenil is a competitive antagonist of benzodiazepines at benzodiazepine receptors and is used to reverse their effects.[15]

Partial agonist action at a receptor

Partial agonists are substances that can be agonists or antagonists at receptors, depending on the endogenous tone of the system upon which they are acting. If the degree of endogenous stimulation of the receptor is low, a partial agonist will tend to act as an agonist; if high, it will end to act as an antagonist.

The anxiolytic buspirone is a partial agonist at 5-HT_{1A} autoreceptors and reduces the firing of 5-hydroxytryptamine (5-HT) neurones by stimulating the auto-receptors.[16]

Actions via second messengers and ion channels

Some drugs act directly on second messenger systems and ion channels, without actions at receptors.

Lithium inhibits enzymes involved in the metabolism of inositol phosphates and may deplete cells of phosphoinositides, which are important as second messengers in neurotransmission.[17] However, other mechanisms have been proposed, including effects on the synthesis, turnover, and functional activity of brain 5-HT and effects on neuronal membrane function by effects on sodium and potassium fluxes via the sodium/potassium pump enzyme.[18]

Altered neurotransmitter storage

Reserpine, which causes depression, inhibits the incorporation of neurotransmitters into presynaptic storage vesicles and thus causes depletion of neurotransmitter stores.

Increased neurotransmitter release

Amphetamines cause increased release of noradrenaline (norepinephrine) and dopamine and have mood-enhancing effects.[19]

Inhibition of neurotransmitter reuptake

Most antidepressants inhibit the reuptake of monoamines into the presynaptic nerve ending after their release.[20] The effects of different antidepressants on monoamine reuptake are listed in Table 6.2.1.2. Reduced reuptake of monoamines occurs immediately, but the full therapeutic effects of antidepressants take some weeks to occur. This is explained by the occurrence of adaptive changes in presynaptic and postsynaptic receptors, including down-regulation of β-adrenoceptors, reduced sensitivity of β- and α_2-adrenoceptors, and increased sensitivity of α_1-adrenoceptors and 5-HT receptors. However, it is not known how these actions are translated into the therapeutic effect. The last of these effects has led to the use of 5-HT autoreceptor antagonists, in the hope of producing a quicker onset of antidepressant action by enhancing 5-HT neurotransmission.[21,22] For example, the partial β-adrenoceptor agonist and 5-HT_{1A} receptor antagonist pindolol hastens the response to SSRIs, although it does not affect the extent of the response.[23] Another strategy involves drugs with several actions, which inhibit more than one reuptake system and are antagonists at neurotransmitter receptors.[24]

Table 6.2.1.2 Effects of antidepressants on monoamine reuptake

Drug	Inhibition of uptake		
	Noradrenaline	5-HT	Dopamine
Tricyclics			
Amitriptyline	+++	++	—
Clomipramine	+++	+++	—
Desipramine	+++	++	—
Imipramine[a]	++[a]	+	—
Nortriptyline	++	+	—
Tetracyclics			
Maprotiline	++	—	—
Mianserin	++	—	—
Phenylethylamines			
Venlafaxine	++	++++	—
Triazolopyridines			
Trazodone	—	+	—
Specific serotonin reuptake inhibitors			
Fluoxetine	—	+++	—
Fluvoxamine	+	++	—
Paroxetine	—	+++	—
Sertraline	+	++++	+

[a] Through its metabolite desipramine.

Altered neurotransmitter metabolism

Inhibitors of monoamine oxidase (MAO) enhance monoamine neurotransmission by irreversibly and non-selectively inhibiting the breakdown of monoamines after their release. Moclobemide[25] is a RIMA, a reversible inhibitor of MAO type A, which metabolizes 5-HT and noradrenaline. Moclobemide is therefore less likely to cause hypertension when taken in combination with amine-containing foods, such as cheese, since this reaction requires inhibition of both MAO type A and MAO type B.

Valproate partly acts by inhibiting GABA transaminase, thus enhancing GABA inhibitory transmission. However, it has other actions: it inhibits GABA reuptake, increases the sensitivity of GABA receptors to GABA, reduces the concentrations of the excitatory neurotransmitter aspartate, and may open potassium channels, thus stabilizing neuronal cell membranes.

Adverse effects of drugs

Unwanted effects of drugs are commonly referred to as toxic effects or side effects. However, these are ambiguous terms, for several reasons:

- Toxic effects occur through exaggeration of the desired pharmacological action of the drug, and therefore occur at doses that are above those usually associated with a therapeutic effect. For example, antipsychotic drugs produce some toxic effects by antagonism at dopamine receptors in the extrapyramidal tracts.
- Toxic effects can also occur through exaggeration of actions other than those that are thought to produce the therapeutic action. Paracetamol toxicity occurs because an active metabolite binds covalently to liver proteins, damaging them.
- Side effects occur either through actions that are unrelated to the desired pharmacological effect or through actions that are related to the desired pharmacological effect but occur in another tissue. Tricyclic antidepressants cause dry mouth, glaucoma, and urinary retention by anticholinergic action. Sildenafil causes colour vision disturbances by inhibiting phosphodiesterase type V in the eye, the action by which it has its therapeutic effect in erectile dysfunction. True side effects are more properly called collateral effects.[26]

It is therefore better to use the terms 'unwanted effects' or 'adverse effects/reactions'. Adverse reactions and adverse effects are identical—the former are seen from the point of view of the patient, the latter from the point of view of the drug.

Adverse drug effects are classified according to the scheme known as DoTS (Dose, Time, and Susceptibility).[26]

Classification according to dose-relatedness

1 *Hypersusceptibility reactions*: Here the dose–response curve for harm is far to the left of the dose–response curve for benefit; hypersusceptibility adverse reactions therefore occur at doses below those that are normally beneficial. Penicillin allergy is an example.

2 *Collateral reactions*: Here the dose–response curve for harm is in a region that is bounded by a curve that is just to the left of the dose–response curve for benefit and one that is just to the right; collateral adverse reactions therefore occur at doses within the range of those that are normally beneficial. They can occur (i) through a pharmacological effect that is distinct from that involved in the beneficial effect (for example an anticholinergic effect of a tricyclic antidepressant) or (ii) through the same pharmacological effect as that associated with the beneficial effect, but in a different tissue (for example colour vision disturbance due to sildenafil).

3 *Toxic reactions*: Here the harm occurs through the same mechanism as benefit (i.e. is on the same dose–response curve) but at doses that are above those that are normally beneficial. An example is serotonin syndrome due to fluoxetine.

Classification according to time-relatedness

Adverse drug reactions can be either time-dependent or time-independent (Table 6.2.1.3).

1 Time-independent reactions can occur at any time during therapy and are generally toxic reactions. They occur when the actual concentration of the drug increases or dose response curve shifts to the left, for whatever reason. An example is digoxin toxicity, which can occur for pharmaceutical reasons (for example administration of the wrong tablets), pharmacokinetic reasons (for example renal insufficiency), or pharmacodynamic reasons (for example hypokalemia). In the first two cases the concentration at the site of action increases and in the last the dose–response curve is shifted to the left.

2 Time-dependent adverse drug reactions are of six types; examples are given in Table 6.2.1.3.
   - Immediate or rapid reactions occur when a drug is given too quickly.
   - First-dose reactions occur only after the first dose of a course.
   - Early reactions occur soon after the first administration; they either wear off with time (early tolerant effects) or persist (early persistent effects).

Table 6.2.1.3 Time-related classification of adverse drug reactions in the DoTS method. (Reproduced from *British Medical Journal*, Aronson, J.K. and Ferner, R.E. (2005), **327**, 1222–5, with permission from BMJ Publishing Group Ltd.)

Type of reaction	Examples	Implications
Time independent		
Due to a change in dose or concentration (pharmaceutical effects)	Toxicity due to increased systemic availability	Beware of changing formulations of some drugs (e.g. modified-release formulations of lithium)
Due to a change in dose or concentration (pharmacokinetic effects)	Lithium toxicity due to renal insufficiency	Forewarn the patient; monitor carefully throughout treatment; alter dosage when pharmacokinetics change (e.g. renal insufficiency); avoid interacting drugs
Occurs without a change in dose (pharmacodynamic effects)	Digitalis toxicity due to hypokalaemia	Forewarn the patient; monitor carefully throughout treatment; avoid precipitating (pharmacodynamic) factors; avoid interacting drugs
Time dependent		
Immediate (due to rapid administration)	Red man syndrome (vancomycin) Hypertension (digitalis) Hypotension (iodipamide)	Administer slowly
First dose [of a course]	Hypotension (Ð1 adrenoceptor antagonists and angiotensin converting enzyme inhibitors) Type I hypersensitivity reactions	Take special precautions for the first dose Careful history taking; if a reaction occurs, avoid re-exposure; counsel the patient
Early tolerant	Adverse reactions that involve tolerance (e.g. nitrate-induced headache)	Monitor during the early stages; give appropriate reassurance; expect adverse effects if strategies to avoid tolerance are adopted
Early persistent	Glucocorticoid-induced diabetes mellitus	Monitor during the early stages and treat appropriately or withhold
Intermediate (risk increases at first, then diminishes)	Venous thromboembolism (classical antipsychotic drugs) Neutropenia (clozapine) Hypersensitivity reactions types II, III, and IV	Monitoring not needed after the high-risk period unless susceptibility changes; withdraw drug if a reaction develops
Late (risk increases with time)	Osteoporosis (glucocorticosteroids) Tardive dyskinesia (dopamine receptor antagonists) Retinopathy (chloroquine) Tissue phospholipid deposition (amiodarone)	Assess baseline function; forewarn the patient; monitor periodically during prolonged treatment
	Withdrawal syndromes: opiates, benzodiazepines, hypertension (clonidine and methyldopa), myocardial infarction (beta-blockers)	Withdraw slowly; forewarn the patient; replace with a longer acting drug if withdrawal is not possible
Delayed	Carcinogenesis (ciclosporin, diethylstilbestrol) Teratogenesis (thalidomide)	Avoid or screen; counsel or forewarn the patient

- Intermediate reactions occur within the first few weeks or months of administration but not thereafter; those who are susceptible will suffer the reaction and those who are not will not (healthy survivors); for example, clozapine causes neutropenia predominantly during the first 24 weeks of therapy—thereafter the risk is small.
- Late reactions occur late in the course of administration, the risk increasing with time; this group includes withdrawal reactions.
- Delayed reactions are seen at some distant time after the initial exposure, even if the drug is withdrawn before the reaction appears.

Classification according to susceptibility factors

The risk of an adverse drug reaction differs among members of an exposed population. For some reactions some individuals are susceptible, others are not—for example, prolonged muscle relaxation due to suxamethonium in people with pseudocholinesterase deficiency. In other cases susceptibility follows a continuous distribution—for example, increasing susceptibility with increasing impairment of renal function. Although reasons for increased susceptibility may be unknown, several types are recognized (Table 6.2.1.4).[27–34] These include:

- genetic variation;
- age;
- sex;
- physiological variation (for example pregnancy, body weight);
- exogenous factors (for example drugs and food);
- diseases (for example renal or hepatic impairment).

More than one susceptibility factor can be present in an individual.

Table 6.2.1.4 Sources of susceptibility to adverse drug reactions

Source of susceptibility*	Examples	Implications
Genetic	Porphyria Suxamethonium sensitivity Malignant hyperthermia CYP isozyme polymorphisms	Screen for abnormalities; avoid specific drugs
Age	Neonates (chloramphenicol[27]) Elderly people (hypnotics[28])	Adjust dosages according to age
Sex	Alcohol intoxication Mefloquine, neuropsychiatric effects[29] Lupus-like syndrome[30]	Use different doses in men and women
Physiology altered	Phenytoin in pregnancy[31]	Alter dosage or avoid
Exogenous factors	Drug interactions Interactions with food (e.g. grapefruit juice with drugs cleared by CYP3A4[32]; see Table 6.1)	Alter dosage or avoid co-administration
Diseases	Renal insufficiency (e.g. lithium[33]) Hepatic cirrhosis (e.g. morphine[34])	Screen for abnormalities; avoid specific drugs; use reduced dosages

*Mnemonic GASPED

Factors that increase the risk of an adverse effect

(a) Pharmaceutical factors

Adverse effects can arise from changes in the pharmaceutical formulation. There is a risk of lithium toxicity or loss of action when one modified-release formulation of lithium is replaced by another.

(b) Pharmacokinetic factors

Changes in the pharmacokinetics of a drug can result in toxic or collateral adverse effects. This most commonly occurs through impaired liver function (Table 6.2.1.1) or renal insufficiency (lithium).

(c) Pharmacodynamic factors

Changes in the sensitivity of a tissue to a drug occur during long-term therapy and can result in adverse effects. Tardive dyskinesia with dopamine receptor antagonists may be related to altered sensitivity of dopamine receptors,[35] although there are problems with this hypothesis, and complex interactions with other neurotransmitters may be involved.[36]

When adaptive changes occur during long-term therapy, sudden withdrawal of the drug can result in rebound reactions. Examples include the typical syndromes that occur after the sudden withdrawal of narcotic analgesics[37,38] or of alcohol (delirium tremens).[39] Sudden withdrawal of barbiturates can cause restlessness, sleeplessness, mental confusion, and convulsions; a similar syndrome, in which anxiety features prominently, can occur after the sudden withdrawal of benzodiazepines.[24,40]

Drug interactions

In a drug interaction one drug alters the effects of another, resulting in increased or decreased effects. Fluvoxamine inhibits the hepatic metabolism of warfarin and increases its anticoagulant effect,[41] whereas carbamazepine reduces the anticoagulant effect of warfarin by increasing its metabolism.[42]

1 Pharmaceutical interactions occur when there is a physicochemical interaction between two compounds in solution. Lists of such incompatibilities are too long to remember. To avoid pharmaceutical interactions, do not combine drugs in an infusion solution and use only sodium chloride 0.9% or glucose 5% in drug infusions.

2 Pharmacokinetic interactions occur when one drug interferes with the disposition of another during absorption, distribution, or elimination. Here are examples that are relevant to psychotropic drugs.

- Sucralfate reduces the absorption of amitriptyline by about 50 per cent;[43] this might be of clinical importance, but drug absorption interactions are not usually important.
- Phenytoin is displaced from protein-binding sites by valproate (which also inhibits its metabolism).[44]
- The metabolism of psychotropic drugs can theoretically be inhibited by drugs that inhibit metabolism via cytochrome P450 in the liver (Table 6.2.1.1). Reports of such interactions appear frequently. Inhibitory drugs include cimetidine, antifungal imidazoles (such as fluconazole), and macrolides (such as erythromycin).
- Non-selective MAO inhibitors, such as tranylcypromine, inhibit the metabolism (by MAO type A) of dietary amines in the gut and metabolism (by MAO type B) of the noradrenaline that they release, resulting in hypertension. Avoid this combination.
- Diuretics inhibit the renal excretion of lithium; alter the dosage of lithium and monitor the serum concentration when starting a diuretic or changing the diuretic dosage.

3 Pharmacodynamic interactions occur when two drugs interact at the same site of action. Alcohol potentiates the actions of all psychotropic drugs; patients taking psychotropic drugs should be warned that even one alcoholic drink can impair their ability to drive or operate machinery. The combination of reuptake inhibitors with non-selective irreversible monoamine oxidase inhibitors can cause the serotonin syndrome,[45] which is

sometimes fatal.[46] The combination of lithium with either SSRIs[47] or the SNRI venlafaxine[48] can also cause the serotonin syndrome. Flumazenil reverses the effects of benzodiazepines—a beneficial pharmacodynamic interaction.

Monitoring drug therapy

If possible, monitor drug therapy by observing the clinical outcome. In psychiatric disorders this is difficult, but it can be done by asking patients or their carers to keep diaries of symptoms.

Next best is to monitor some pharmacological effect of the drug, in the way that one measures the international normalized ratio (INR) in patients taking warfarin (thus measuring the effect of the drug on the blood, being unable to monitor its clinical effect of preventing pulmonary embolism). However, there are no comparable routine tests available for monitoring the pharmacological effects of psychotropic drugs.

Because of these difficulties one falls back on measurements of serum concentrations of some drugs. The assumptions in doing this are that the serum concentration reflects the concentration at the site of action and that there is a concentration–effect relationship. Few psychoactive drugs can be monitored in this way, the principal ones being lithium and phenytoin. Some advocate monitoring treatment with some tricyclic antidepressants,[49] but this use is controversial.

Serum concentration measurement can be used to individualize therapy in the early stages of treatment or when the dosage is being changed, to check adherence to therapy, to help diagnose toxicity, and to monitor the effects of drug interactions.

Carbamazepine

The target plasma carbamazepine concentration range is 17–42 mmol/l, but plasma concentrations do not correlate well with effect, since it has an active metabolite, oxcarbazepine; higher concentrations are associated with an increased risk of toxicity.[50] Carbamazepine induces its own metabolism, and its half-life is therefore shortened during long-term therapy. So, after an initial apparent steady state has been reached 3 or 4 days after starting therapy, a new steady state occurs at a lower concentration a few weeks later, and the dose may need to be increased at that time. Blood samples should be taken immediately before a dose.

Lithium

The target serum lithium concentration is 0.4–1.0 mmol/l. Concentrations above 1 mmol/l are associated with an increased risk of toxicity. Take blood samples at a standard time—12 h after the last dose. In my view, routine monitoring is unnecessary, and the serum lithium concentration should be measured at times when toxicity is most likely, for example in patients with changing renal function or with acute alterations in electrolyte balance.[51,52] Many psychiatrists prefer to measure the serum lithium concentration routinely at, say, 3- or 6-monthly intervals. However, although regular monitoring may emphasize the dangers to the patient and give an opportunity for a consultation, it is no substitute for proper monitoring at the appropriate times.

Valproate

The target serum valproate concentration range is 40–80 μmol/1, although this range is based on its use in epilepsy rather than bipolar affective disorder;[53] higher concentrations are associated with an increased risk of toxicity. The blood sample can be taken at any time after the last dose.

Further information

Aronson, J.K. (ed.) (2006). *Meyler's side effects of drugs. The international encyclopedia of adverse drug reactions and interactions* (15th edn). Elsevier, New York.

Aronson, J.K., Hardman, M., and Reynolds, D.J.M. (1993). *ABC of monitoring drug therapy*. BMJ Publishing, London.

Association of British Pharmaceutical Industries. *ABPI compendium of data sheets and summaries of product characteristics*. Datapharm Publications, London (published annually).

Baxter, K. (ed.) (2006). *Stockley's drug interactions: a source book of interactions, their mechanisms, clinical importance and management* (7th edn). Pharmaceutical Press, London.

Bennett, P.N. (ed.) (1996). *Drugs and human lactation* (2nd edn). Elsevier, Amsterdam.

Denham, M.J. and George, C.F. (eds.) (1990). Drugs in old age. *British Medical Bulletin*, **46**.

Dollery, C. (ed.) (1999). *Therapeutic drugs* (2nd edn). Churchill Livingstone, Edinburgh.

Grahame-Smith, D.G. and Aronson, J.K. (2002). *Oxford textbook of clinical pharmacology* (3rd edn). Oxford University Press, Oxford.

Gilstrap, L.C. and Little, B.B. (eds.) (1998). *Drugs and pregnancy* (2nd edn). Chapman & Hall, New York.

Joint Formulary Committee. *British national formulary*. British Medical Association and The Pharmaceutical Society of Great Britain, London (published every 6 months).

Paediatric Formulary Committee. *British national formulary for children*. British Medical Association and The Pharmaceutical Society of Great Britain, London (published annually).

References

1. Amdisen, A. (1969). Variation of serum lithium concentration during the day in relation to treatment control, absorptive side effects and the use of slow-release tablets. *Acta Psychiatrica Scandinavica*, **207**, 55–7.
2. Malone, F.D. and D'Alton, M.E. (1997). Drugs in pregnancy: anticonvulsants. *Seminars in Perinatology*, **21**, 114–23.
3. Rodriguez-Pinilla, E., Arroyo, I., Fondevilla, J., *et al.* (2000). Prenatal exposure to valproic acid during pregnancy and limb deficiencies: a case-control study. *American Journal of Medical Genetics*, **90**, 376–81.
4. Jefferson, J.W. (2006). Lithium. In *Meyler's side effects of drugs. The international encyclopedia of adverse drug reactions and interactions* (15th edn) (ed. J.K. Aronson), pp. 2073–116. Elsevier, Amsterdam.
5. Anonymous (1999). Lithium. In *Therapeutic drugs* (2nd edn), Vol. 2 (ed. C. Dollery), pp. L71–5. Churchill Livingstone, Edinburgh.
6. Chisholm, C.A. and Kuller, J.A. (1997). A guide to the safety of CNS–active agents during breastfeeding. *Drug Safety*, **17**, 127–42.
7. Pentikainen, P.J., Neuvonen, P.J., and Jostell, K.G. (1980). Pharmacokinetics of chlormethiazole in healthy volunteers and patients with cirrhosis of the liver. *European Journal of Clinical Pharmacology*, **17**, 275–84.
8. Caccia, S. (1998). Metabolism of the newer antidepressants. An overview of the pharmacological and pharmacokinetic implications. *Clinical Pharmacokinetics*, **34**, 281–302.
9. Norman, T.R., Walker, R.G., and Burrows, G.D. (1984). Renal function related changes in lithium kinetics. *Clinical Pharmacokinetics*, **9**, 349–53.

10. Korpi, E.R., Mattila, M.J., Wisden, W., *et al.* (1997). GABA(A) receptor subtypes: clinical efficacy and selectivity of benzodiazepine site ligands. *Annals of Medicine*, **29**, 275–82.
11. Wagner, J., Wagner, M.L., and Hening, W.A. (1998). Beyond benzodiazepines: alternative pharmacologic agents for the treatment of insomnia. *Annals of Pharmacotherapy*, **32**, 680–91.
12. Cross, A.J., Stirling, J.M., Robinson, T.N., *et al.* (1989). The modulation by chlormethiazole of the $GABA_A$-receptor complex in rat brain. *British Journal of Pharmacology*, **98**, 284–90.
13. Evans, R.W. and Lipton R.B. (2001). Topics in migraine management: a survey of headache specialists highlights some controversies. *Neurologic Clinics*, **19**, 1–21.
14. Seeman, P. and Tallerico, T. (1998). Antipsychotic drugs which elicit little or no parkinsonism bind more loosely than dopamine to brain D_2 receptors, yet occupy high levels of these receptors. *Molecular Psychiatry*, **3**, 123–34.
15. Weinbroum, A.A., Flaishon, R., Sorkine, P., *et al.* (1997). A risk benefit assessment of flumazenil in the management of benzodiazepine overdose. *Drug Safety*, **17**, 181–96.
16. Gardner, C.R. (1988). Potential use of drugs modulating 5HT activity in the treatment of anxiety. *General Pharmacology*, **19**, 347–56.
17. Harwood, A.J. (2005). Lithium and bipolar mood disorder: the inositol-depletion hypothesis revisited. *Molecular Psychiatry*, **10**, 117–26.
18. Marmol, F. (2006). Litio: 55 anos de historia en el tratamiento del trastorno bipolar. *Medica Clinica* (*Barcelona*), **127**, 189–95.
19. Fleckenstein, A.E., Volz, T.J., Riddle. E.L., *et al.* (2007). New insights into the mechanism of action of amphetamines. *Annual Reviews of Pharmacology and Toxicology*, **47**, 681–98.
20. Stahl, S.M. (1998). Basic psychopharmacology of antidepressants, part 1: antidepressants have seven distinct mechanisms of action. *The Journal of Clinical Psychiatry*, **59**(Suppl. 4), 5–14.
21. Blier, P. and Bergeron, R. (1997). Early onset of therapeutic action in depression and greater efficacy of antidepressant treatments: are they related? *International Clinical Psychopharmacology*, **12**(Suppl. 3), S21–8.
22. Sussman, N. and Joffe, R.T. (1998). Antidepressant augmentation: conclusions and recommendations. *The Journal of Clinical Psychiatry*, **59**(Suppl. 5), 70–3.
23. Artigas, F., Adell, A., and Celada, P. (2006). Pindolol augmentation of antidepressant response. *Current Drug Targets*, **7**, 139–47.
24. Sambunaris, A., Hesselink, J.K., Pinder, R., *et al.* (1997). Development of new antidepressants. *The Journal of Clinical Psychiatry*, **58**(Suppl. 6), 40–53.
25. Fulton, B. and Benfield, P. (1996). Moclobemide. An update of its pharmacological properties and therapeutic use. *Drugs*, **52**, 450–74 (erratum 869).
26. Aronson, J.K. and Ferner, R.E. (2003). Joining the DoTS. New approach to classifying adverse drug reactions. *British Medical Journal*, **327**, 1222–5.
27. Mulhall, A., de Louvois, J., and Hurley, R. (1983). Chloramphenicol toxicity in neonates: its incidence and prevention. *British Medical Journal*, **287**, 1424–7.
28. Greenblatt, D.J., Divoll, M., Harmatz, J.S., *et al.* (1981). Kinetics and clinical effects of flurazepam in young and elderly noninsomniacs. *Clinical Pharmacology and Therapeutics*, **30**, 475–86.
29. Schwartz, E., Potasman, I., Rotenberg, M., *et al.* (2001). Serious adverse events of mefloquine in relation to blood level and gender. *American Journal of Tropical Medicine and Hygiene*, **65**, 189–92.
30. Batchelor, J.R., Welsh, K.I., Tinoco, R.M., *et al.* (1980). Hydralazine-induced systemic lupus erythematosus: influence of HLA-DR and sex on susceptibility. *Lancet*, **1**, 1107–9.
31. Dickinson, R.G., Hooper, W.D., Wood, B., *et al.* (1989). The effect of pregnancy in humans on the pharmacokinetics of stable isotope labelled phenytoin. *British Journal of Clinical Pharmacology*, **28**, 17–27.
32. Aronson, J.K. (2001). Forbidden fruit. *Nature Medicine*, **7**, 7–8.
33. Clericetti, N. and Beretta-Piccoli, C. (1991). Lithium clearance in patients with chronic renal diseases. *Clinical Nephrology*, **36**, 281–9.
34. Hasselstrom, J., Eriksson, S., Persson, A., *et al.* (1990). The metabolism and bioavailability of morphine in patients with severe liver cirrhosis. *British Journal of Clinical Pharmacology*, **29**, 289–97.
35. Jenner, P. and Marsden, C.D. (1987). Chronic pharmacological manipulation of dopamine receptors in brain. *Neuropharmacology*, **26**, 931–40.
36. Gill, H.S., DeVane, C.L., and Risch, S.C. (1997). Extrapyramidal symptoms associated with cyclic antidepressant treatment: a review of the literature and consolidating hypotheses. *Journal of Clinical Psychopharmacology*, **17**, 377–89.
37. Puntillo, K., Casella, V., and Reid, M. (1997). Opioid and benzodiazepine tolerance and dependence: application of theory to critical care practice. *Heart & Lung*, **26**, 317–24.
38. O'Connor, P.G. and Kosten, T.R. (1998). Rapid and ultrarapid opioid detoxification techniques. *The Journal of the American Medical Association*, **279**, 229–34.
39. Miller, N.S. (1995). Pharmacotherapy in alcoholism. *Journal of Addictive Diseases*, **14**, 23–46.
40. Ashton, H. (1994). The treatment of benzodiazepine dependence. *Addiction*, **89**, 1535–41.
41. Perucca, E., Gatti, G., and Spina, E. (1994). Clinical pharmacokinetics of fluvoxamine. *Clinical Pharmacokinetics*, **27**, 175–90.
42. Cropp, J.S. and Bussey, H.I. (1997). A review of enzyme induction of warfarin metabolism with recommendations for patient management. *Pharmacotherapy*, **17**, 917–28.
43. Ryan, R., Carlson, J., and Farris, F. (1986). Effect of sucralfate on the absorption and disposition of amitriptyline in humans. *Federation Proceedings*, **45**, 205.
44. Suzuki, Y., Nagai, T., Mano, T., *et al.* (1995). Interaction between valproate formulation and phenytoin concentrations. *European Journal of Clinical Pharmacology*, **48**, 61–3.
45. Hilton, S.E., Maradit, H., and Moller, H.J. (1997). Serotonin syndrome and drug combinations: focus on MAOI and RIMA. *European Archives of Psychiatry and Clinical Neuroscience*, **247**, 113–9.
46. Keltner, N. and Harris, C.P. (1994). Serotonin syndrome: a case of fatal SSRI/MAOI interaction. *Perspectives in Psychiatric Care*, **30**, 26–31.
47. Sobanski, T., Bagli, M., Laux, G., *et al.* (1997). Serotonin syndrome after lithium add-on medication to paroxetine. *Pharmacopsychiatry*, **30**, 106–7.
48. Adan-Manes, J., Novalbos, J., Lopez-Rodriguez, R., *et al.* (2006). Lithium and venlafaxine interaction: a case of serotonin syndrome. *Journal of Clinical Pharmacy and Therapy*, **31**, 397–400.
49. Isacsson, G., Bergman, U., Wasserman, D., *et al.* (1996). The use of antidepressants and therapeutic drug monitoring by general practitioners and psychiatrists: findings from a questionnaire survey in two Swedish areas. *Annals of Clinical Psychiatry*, **8**, 153–60.
50. Vasudev, K., Goswami, U., and Kohli, K. (2000). Carbamazepine and valproate monotherapy: feasibility, relative safety and efficacy, and therapeutic drug monitoring in manic disorder. *Psychopharmacology* (*Berlin*), **150**, 15–23.
51. Aronson, J.K. and Reynolds, D.J. (1992). ABC of monitoring drug therapy. Lithium. *British Medical Journal*, **305**, 1273–4.
52. Glasziou, P. and Aronson, J.K. (2008). An introduction to monitoring therapeutic interventions in clinical practice. In *Evidence-based medical monitoring: from principles to practice* (eds. P.P. Glasziou, L. Irwig, and J.K. Aronson), Chap. 1. Wiley-Blackwell Publications, Oxford.
53. Fleming, J. and Chetty, M. (2006). Therapeutic monitoring of valproate in psychiatry: how far have we progressed? *Clinical Neuropharmacology*, **29**, 350–60.

6.2.2 Anxiolytics and hypnotics

Malcolm Lader

Introduction

Anxiety is a commonly experienced emotion that becomes a clinical disorder when it is too severe, too protracted, or too pervasive for the subject to bear. Insomnia is a failure to experience satisfying sleep, together with a feeling of tiredness during the day. Many compounds, the anxiolytics and hypnotics, are used to treat these conditions, but the two groups of drugs overlap.

The classical antianxiety drugs (anxiolytics) are alcohol, the opioids, and the barbiturates. For the past 45 years, the benzodiazepines, such as diazepam and lorazepam, have dominated the field. They are effective anxiolytics in the short term but their long-term efficacy remains in dispute. Their disadvantages include cognitive and psychomotor impairment, paradoxical reactions, tolerance, and dependence, and they are major drugs of abuse.

Other anxiolytics act on the 5-hydroxytryptamine (5-HT; serotonin) systems of the brain and include buspirone and the selective serotonin reuptake inhibitors (SSRIs). Newer compounds are still being introduced that lie outside these groups.

The use of benzodiazepine and benzodiazepine-like hypnotics, by contrast, continues apace. Some switching to the shorter-acting benzodiazepines has occurred, together with the introduction of the 'z-compounds', zopiclone, zolpidem, and zaleplon. These drugs tend to have fewer residual effects the next day than the benzodiazepines, and are claimed to be less likely to induce rebound and dependence than equivalent benzodiazepines. Particular care is needed in prescribing such hypnotics to the elderly.

The rational use of both anxiolytics and hypnotics requires minimal dosage, short durations of use, and simultaneous exploitation of non-pharmacological methods.

Definitions

'Sedative' originally meant a substance that has the property of allaying anxiety. However, it has now come to denote feelings of drowsiness or torpor. This state was originally called 'oversedation', and was often noted with the barbiturates and other older drugs such as chloral. Next, the term 'tranquillizer' was introduced 40 or more years ago in an attempt to distinguish between the older sedatives and the newer drugs, supposedly non-sedative, such as the benzodiazepines. But this distinction is artificial as, apart from safety in overdosage, the benzodiazepines closely resemble the barbiturates in pharmacological and clinical properties. The term 'anxiolytic' is now generally favoured.

Anxiolytic drugs

Anxiety-allaying drugs have been used for thousands of years, dating back to the discovery that, among its psychotropic properties, alcohol could induce sedation. The nineteenth century saw the development of inorganic and, later, organic chemical compounds. Bromides were introduced as sedatives and became widely used despite their poor effectiveness, toxicity, and potential abuse. Organic chemists in the second half of the nineteenth century introduced sedatives such as chloral and paraldehyde.

The first barbiturate was introduced over a 100 years ago. This group of drugs is divisible into the ultrashort-acting (e.g. anaesthetic-induction agents such as thiopentone and methohexital), short-acting (e.g. secobarbital), medium-acting (e.g. butobarbital), and long-acting (e.g. phenobarbital) barbiturates. Most of the rest are of medium duration with half-lives of 16 h or so. The disadvantages of the barbiturates include drowsiness, tolerance to their effects, dangers of overdose, and possible physical and psychological dependence with severe withdrawal syndromes.[1] Meprobamate was introduced as the first of the 'tranquillizers', but its advantages over the barbiturates proved minimal.

The benzodiazepines were first synthesized in the 1930s, but not developed until 2 years later. The prototype, chlordiazepoxide, was evaluated in the clinic, found effective, and soon introduced into medical practice. More than 1000 benzodiazepines and related compounds have been synthesized, including diazepam, the most widely used of all. Anxiolytic and hypnotic, as well as muscle-relaxant and anticonvulsant properties are licensed indications. However, the distinction between anxiolytic and hypnotic uses often seems to owe more to commercial expediency than to scientific rationale; some compounds, such as lorazepam, are marketed for both indications.

The benzodiazepines

The main reason for the original popularity of the benzodiazepines was the perceived safety in overdose compared with the quite marked toxicity of the barbiturates. In turn, concern has mounted concerning the benzodiazepines.[2] These drugs are widely prescribed by many physicians for patients with emotional problems, circulatory disorders, tension headaches, and pains in the chest and back as well as digestive disorders, all with the common symptom of anxiety. This widespread use, even overuse and the induction of dependence even at normal therapeutic dose has led to official injunctions for greater caution in prescribing.

Pharmacokinetics

Two aspects of the pharmacokinetics of the benzodiazepines are relevant to the prescriber—speed of onset of action and the duration of that action. The speed of onset depends on the mode of administration and the penetration time to the brain. Given by mouth, most benzodiazepines are rapidly absorbed and exert a prompt anxiolytic effect, for instance in panic states. Diazepam and lorazepam are prime examples. Although temazepam enters the brain more rapidly than, say, oxazepam, it still takes an appreciable time to induce sleep. The redistribution phase can be pronounced and will then largely determine the duration of effect of single doses of benzodiazepines such as diazepam and flunitrazepam.

The metabolic half-lives of the benzodiazepines also vary greatly. *N*-desmethyldiazepam (nordiazepam) is the major and active metabolite of diazepam and several other benzodiazepines. It has a long half-life, about 60 h, and accumulates over the first month of treatment. Metabolism of these drugs is even slower in the elderly and in patients with liver damage.

Benzodiazepines with a 3-hydroxyl grouping, such as lorazepam, oxazepam, and temazepam, have half-lives averaging 12 h or less. Liver damage has to be severe before the metabolism of these drugs

is affected. Alprazolam is a triazolobenzodiazepine with a half-life of 9 to 16 h, and with hydroxy metabolites of low biological activity. Both chlordiazepoxide and diazepam are absorbed erratically after intramuscular injection. Lorazepam, however, is well absorbed after intramuscular injection.

Basic pharmacology

The benzodiazepines potentiate the widespread inhibitory neurotransmitter γ-amino butyric acid (GABA). Benzodiazepines do not act directly on GABA receptors but have their own receptors. Because of this widespread inhibitory effect, benzodiazepines alter the turnover of neurotransmitters such as norepinephrine and serotonin. The main sites of action of the benzodiazepines are in the spinal cord where muscle-relaxant effects are mediated, the brainstem (perhaps accounting for their anticonvulsant properties), the cerebellum (causing ataxia), and the limbic and cortical areas involved in the organization of emotional experience and behaviour.

Clinical pharmacology

The depressant effects of single therapeutic doses of a benzodiazepine can usually be readily detected. However, lower doses may fail to impair psychological functioning and subjective effects are usually absent. In the clinical context with anxious patients and with repeated higher doses, sustained impairment of functioning is more difficult to demonstrate. Some studies have shown decrements in performance after the first dose, but improvements in functioning, in comparison to predrug levels, may become apparent by the end of a week of repeated usage. This suggests that the well-known impairment of performance produced by pathologically high levels of anxiety is first worsened by the sedative effects. Then as the antianxiety effects build-up, the patient's psychological functions may improve.

A second mechanism concerns tolerance, which reflects several biochemical mechanisms including alteration in benzodiazepine-receptor type. Patients who have a high alcohol intake are tolerant to benzodiazepines.

The benzodiazepines have marked and selective effects on memory by interfering with episodic memory, that is to say the system concerned with remembering personal experiences.[3] This effect seems independent of any sedation or attentional impairment. Alcohol adds to the cognitive impairment induced by the benzodiazepines but does not necessarily potentiate it.

The dependence potential of benzodiazepines is seen in drug-preference studies, but these drugs are much less preferred than say the amphetamines. Differences among benzodiazepines have been documented; for example, oxazepam seems to have less abuse liability than diazepam.

The largest gap in our knowledge of these drugs is on their long-term usage, which has been evaluated in relatively few studies.[4] Thus, it is still largely unclear whether therapeutic effects are maintained in most patients for longer than a few weeks and when dependence supervenes in the minority of patients who encounter problems on protracted usage.

Hypnotic drugs

The main groups of drugs used in the modern treatment of insomnia are the benzodiazepines, and the newer compounds, zopiclone, eszopiclone, zolpidem, and zaleplon. The pharmacology of these benzodiazepines is essentially the same as that of the anxiolytic compounds.

Nitrazepam is a long-acting benzodiazepine with an elimination half-life ranging between 25 and 35 h, but it is longer in the elderly. Because of this, it is likely to produce residual effects and to accumulate. Flunitrazepam is more potent, but is somewhat shorter acting with a half-life of 10 to 20 h. It has a rapid redistribution phase, which can result in a short duration of intense action. It has earned an undeserved reputation as the 'date-rape' drug. Flurazepam is still widely used in the United States. It has a very long-acting metabolite, which can produce psychological impairment on regular dosage, especially in the elderly. Of the intermediate-acting compounds, temazepam has a half-life of 10 to 15 h, without active metabolites. At modest dose (10–15 mg daily), it results in few residual effects and is fairly well tolerated by the elderly. Major problems with abuse have limited its popularity, but it is still widely prescribed worldwide. Lormetazepam is slightly shorter acting, loprazolam has a fairly short half-life, but its absorption may be slow and erratic.

Triazolam is the archetypal short-acting benzodiazepine, with a mean half-life of around 3 to 4 h, and no clinically significant metabolites. Daytime sedation is seen after high doses (0.5 mg daily), but not usually with lower ones. These higher doses have also been associated with an increased incidence of anterograde amnesia and unusual behaviours, including depressive reactions and hostility.

Zopiclone is a cyclopyrrolone derivative believed to bind close to, but not exactly at, the benzodiazepine receptor. It has a half-life of about 5 h in younger subjects and about 8 h in the elderly. Its sedative and hypnotic effects are similar to those of the benzodiazepines, but its side effect profile is generally superior with fewer central nervous system effects such as oversedation, confusion, and memory impairment. Rebound and withdrawal problems also seem to be less.

Eszopiclone is the S-enantiomer of zopiclone, which is a racemic mixture. It is licensed for the long-term treatment of insomnia in the United States, following successful clinical trials.

Zolpidem is an imidazopyridine compound that binds selectively to one subtype of the benzodiazepine receptor. It is rapidly absorbed and has a short elimination half-life of 0.7 to 3.5 h (mean 2.4 h). It decreases sleep-onset latency but has less consistent effects on total sleep time.[5] Residual effects are minimal, as are memory disturbances. Rebound and withdrawal are uncommon but have been documented.

Zaleplon is also a selective compound with a very short half-life averaging only 1 h. It shortens sleep onset without usually prolonging total sleep time. Residual effects are absent, and memory is minimally disturbed.

Clinical effects of anxiolytics

Although the usual licensed indications are generalized anxiety and panic disorder,[6,7] the main practical application of the benzodiazepines is to aid in the symptomatic management of anxiety and stress-related conditions.[8] These indications are often so wide as to be difficult to define in terms of recognized disorders. Instead the symptoms of anxiety, in whatever context, are the main indication.

Thousands of comparative trials among the benzodiazepines have been carried out, but few differences with respect to risk–benefit ratios have been found.

Antianxiety medications are difficult to assess. Anxiety disorders are very varied in their natural history; some resolve over a few weeks, whereas others become chronic for no apparent reason, with subsequent acute-on-chronic exacerbations. The patients with chronic, severe unresponsive illnesses tend to be referred to psychiatric outpatient departments. Uncontrolled observations on family practice patients will give a more encouraging impression of antianxiety drugs than will assessment of the more chronic patients attending psychiatric clinics. Even in the latter type of patient, useful symptomatic relief is often obtained without complete resolution of the illness.

Drugs such as diazepam have a long elimination half-life so that once daily or nightly dosage is sufficient. Nevertheless, many patients prefer to take a divided dosage during the day, often claiming that they can detect further antianxiety activity after each dose and are thereby reassured. For episodic anxiety, shorter-acting compounds such as lorazepam can be used, taken 30 min or so before entering the anxiety-provoking situation. If the panic has already started, lorazepam can still be given and will exert a fairly prompt action. Lorazepam is also invaluable in the emergency management of the acutely anxious and disturbed psychotic patient.

Antipanic actions have been claimed for the benzodiazepines, in particular alprazolam, acting to prevent the episodes rather than aborting them. However, although suppression of the panic attack is often quite effective, relapse, and even rebound may occur when the benzodiazepine is discontinued, even if it is tapered off.[9] Because of this SSRI antidepressants are generally preferred.[7]

The short-acting benzodiazepines are also used as adjuncts to relaxation therapy, preoperative medication, and deep sedation for minor operative procedures such as dentistry. The drugs render the patient calm, conscious, and cooperative, with often total anterograde amnesia for the operation.

Unwanted effects

The commonest unwanted effects of the anxiolytic benzodiazepines are tiredness, drowsiness, and torpor, features of 'oversedation'. The effects are dose and time related, being maximal within the first 2 h after large doses. Drowsiness is most noticeable during the first week of treatment, after which it largely disappears probably due to a true tolerance effect. Patients should be warned of the potential side effects of any prescribed benzodiazepine and the initial dosage should be cautious. Both psychomotor skills and intellectual and cognitive skills are affected. In particular, patients should be advised not to drive during the initial adjustment of dosage. Important decisions should be deferred during this period because judgement may be clouded.

Benzodiazepines have major effects on cognitive function in long-term users. A meta-analysis of 13 research studies revealed impairments across all cognitive categories examined.[4] The drugs differ in their ability to produce memory deficits, with lorazepam being especially powerful.[3] However, most benzodiazepines can cause problems, especially in higher dose and in the elderly.

Psychomotor performance is also affected, with elderly drivers particularly at risk. As with other depressant drugs, potentiation of the effects of alcohol can occur. Patients must be warned not to drink alcohol when taking benzodiazepines, either chronically or intermittently. Patients taking benzodiazepines may develop paradoxical behavioural responses such as uncontrollable weeping, increased aggression and hostility, and acute rage reactions or uncharacteristic criminal behaviour such as shoplifting.[10] This phenomenon is by no means confined to the benzodiazepines; alcohol is a cardinal example of a drug whose use may lead to excessive violence or criminal behaviour. Paradoxical reactions, including the release of anxiety or hostility, are most common during the initial week of treatment, and usually resolve spontaneously or respond to dose adjustment. Reports of the induction of depression by the benzodiazepines in patients with apparent generalized anxiety disorder are probably the result of an initial misdiagnosis and a failure to detect the underlying depression.

Other unwanted effects include respiratory depression, excessive weight gain, skin rash, impairment of sexual function, menstrual irregularities, and rarely, blood dyscrasias. The use of benzodiazepines in pregnancy is generally regarded as reasonably safe. However, benzodiazepines pass readily into the foetus and can produce respiratory depression in the neonate. Finally, benzodiazepines pass into the mother's milk and can over sedate the baby, so breastfeeding should be discouraged if benzodiazepines are prescribed, especially in high dose.

Overdosage

Overdosage with benzodiazepines is common; deaths are not. Although fatal-overdose statistics contain deaths ascribed to benzodiazepines alone,[11] many such attributions are suspect. Only in children and the physically frail, especially those with respiratory illness, are the benzodiazepines on their own hazardous. However, they can markedly potentiate other central nervous system depressant drugs such as alcohol. Typically, the person falls into a deep sleep but can be roused the administration of the benzodiazepine antagonist, flumazenil.

Tolerance and dependence

If tolerance occurred regularly, then escalation of dosage would be the norm. This does occur with the benzodiazepines, but is fairly uncommon. Escalation of dose is often stepwise, with each increment following a temporary deterioration in psychosocial circumstances. Most patients later reduce the dose as the stress resolves, but others continue the higher dose to which they presumably have developed some tolerance.

Tolerance to the clinical effects in patients maintaining moderate doses of benzodiazepines is now generally accepted.[12] Few controlled observations concern the long-term efficacy of antianxiety compounds in chronically anxious patients. If medication is withdrawn, the original symptoms may reappear. This is taken as evidence that therapeutic benefit still continues. However, it may reflect 'rebound' rather than long-term clinical benefit. Undoubtedly, many chronically anxious patients are helped by their treatment with benzodiazepines, but this raises the question as to the frequency of psychological and physical dependence on these drugs. Dependence is easily demonstrable in those patients who have attained high doses. Rebound and withdrawal symptoms after the long-acting benzodiazepine diazepam are not usually apparent until about 5 to 10 days after discontinuation. It is much shorter in patients discontinuing the shorter-acting benzodiazepines (2–4 days). The mildest symptoms and signs are anxiety, tension, apprehension, dizziness, tremulousness, insomnia, and

anorexia. More severe physical dependence is shown by the withdrawal symptoms of nausea and vomiting, severe tremor, muscle weakness, postural hypotension, and tachycardia. Occasionally, hyperthermia, muscle twitches, convulsions, and confusional psychoses may develop.

After normal dose usage, perceptual changes can be particularly troublesome.[13] The proportion of patients taking benzodiazepines chronically who experience withdrawal symptoms on discontinuing medication ranges between 27 and 45 per cent, depending on the criteria used. Sometimes the withdrawal reactions seem very prolonged[14] or depression may supervene.

Management of withdrawal

It is widely accepted that the most appropriate way to manage patients withdrawing from benzodiazepines is to taper the dose gradually, because the severe symptoms of withdrawal, such as epileptic fits and confusional episodes, are more likely to follow abrupt than gradual withdrawal. Views differ as to the rate of withdrawal. Detailed guidelines,[15,16] based on a consensus view in the United Kingdom, recommend minimal intervention first, usually by a general practitioner (Fig. 6.2.2.1). This may comprise a letter to the long-term user, or an interview on a routine visit, with advice to taper the medication. More active intervention involves careful assessment, education, and then the adoption, with the patient's agreement, of a timetable of about 6 to 8 weeks for withdrawal. Some agencies suggest a month of tapering for every year of benzodiazepine use, but this may result in patients becoming preoccupied with their symptoms. One strategy is to try a fairly brisk withdrawal, say over 6 to 8 weeks, and only resort to more gradual tapering if the symptoms become intolerable. Another ploy is first to substitute a long-acting for a short-acting benzodiazepine, say 10 mg of diazepam for 1 mg of lorazepam, and then to taper off the diazepam later.

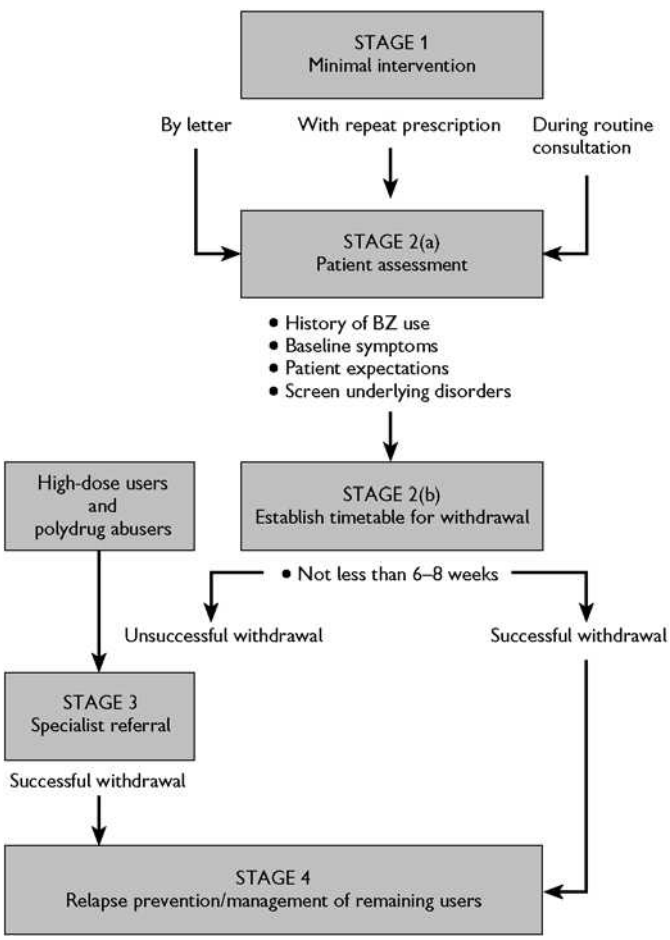


Fig. 6.2.2.1 Guidelines for withdrawal from benzodiazepines (BZ). (Reproduced with permission from J. Russell and M. Lader (eds.), *Guidelines for the prevention and treatment of benzodiazepine dependence*, Mental Health Foundation, London. Copyright 1993, Mental Health Foundation, London.)

Patients must be carefully followed up as a depressive illness is not uncommon and may need vigorous treatment. Such an illness may be reactive to the stress of withdrawal or be a recurrence of an earlier affective episode.

Other drugs have been advocated, but most patients are loath to substitute yet another medication. Depressed patients should have the depression treated before attempting withdrawal. Based on evidence from animal studies, fairly large single doses of flumazenil have been tried with some success.

Psychological support is essential, with the doctor or a practice nurse maintaining close contact with the patient during withdrawal. The physician should show clearly that he understands the problems of withdrawal in order to capture the confidence of the patient. He or she must recognize that patients frequently incubate numerous misconceptions and negative expectations about tranquillizers and withdrawal. These must be elicited, identified, discussed, and corrected.

Cognitive behavioural treatment is currently favoured and is often effective if administered by an experienced professional. Relaxation treatment and training in anxiety management skills in the framework of group therapy can boast of only moderate effectiveness.

Other anxiolytics

Benzodiazepine-receptor partial agonists

The disadvantages of the benzodiazepines include sedation, psychomotor and cognitive impairment, and withdrawal symptoms after long-term use. Increased understanding of benzodiazepine-receptor mechanisms suggested that compounds might be developed which are partial agonists and/or selective to some subtypes of receptor.[17] Such compounds would be less efficacious than full agonists but might have better adverse-effect profiles and less dependence potential, that is superior risk–benefit ratios. Their promise has not been fulfilled, as the risk–benefit ratios of these compounds do not seem superior to those of the full agonists.

5-HT_{1A} partial agonists

These drugs have a complex pharmacology. The first, buspirone, was licensed in many countries some years ago. These drugs suppress activity in presynaptic serotonergic neurones, diminishing serotonin activity, and leading on to down-regulation of 5-HT_2 and perhaps other 5-HT receptors. Buspirone is much less sedative than the benzodiazepines and causes little or no psychomotor or cognitive impairment, nor does it potentiate the effects of alcohol. In formal clinical trials, buspirone was equi-effective and

equipotent to diazepam, but patients taking buspirone improve more slowly.[18] The side effects of buspirone include headache, dizziness, and nausea. Discontinuation is not accompanied by either rebound or withdrawal.

Other 5-HT_{1A} partial agonists have been developed mainly as potential antidepressants and antianxiety agents, but few have been marketed, mainly because of disappointing efficacy.

Pregabalin

This compound is a structural analogue of GABA although it is not active at GABA receptors.[19] It binds with high affinity to an auxiliary α_2-δ subunit protein of voltage-gated calcium channels in the CNS: it acts as a presynaptic modulator of the excessive release of neurotransmitters in hyperexcited neurones. It is predominantly excreted unchanged in the urine. It was initially developed for use as an adjunctive treatment in epilepsy and neuropathic pain. Several clinical trials in GAD have shown it to have efficacy akin to those for benzodiazepines and venlafaxine.[20] It has a rapid onset of action and was effective in preventing relapse over 34 weeks. Tolerance was good during dosage escalation to the usual dose of 150–600 mg/day, mild dizziness and somnolence being the usual adverse effects. No clinically significant withdrawal was seen after tapering. It is approved for the treatment of GAD in Europe, and is a significant introduction.

Antiepileptic drugs

There is a long history of the use of many of these compounds in anxiety disorders,[21] but they are not routine choices.

Antipsychotic drugs

Phenothiazines, such as chlorpromazine and trifluoperazine, and a range of other antipsychotic drugs have been advocated for treating anxiety. The dosage recommended is quite low, typically less than half the initial antipsychotic dose used in psychotic patients. Sometimes, even at this dosage, the antipsychotic drug is not well tolerated by the anxious patient because unwanted autonomic effects, such as dry mouth and dizziness, too closely resemble the symptoms of anxiety. Even more unwelcome are extrapyramidal symptoms such as restlessness (mild akathisia) and parkinsonism, although at the low doses advocated such unwanted effects are uncommon. There may even be a risk of tardive dyskinesia. The chief advantage of this medication is that dependence is virtually unknown, so the main indication for their use is in patients with histories of dependence on other central nervous system depressant drugs such as alcohol or barbiturates.

Antidepressants

Several of these drugs, such as amitriptyline, doxepin, and trazodone, have useful secondary sedative properties. They are widely prescribed for depressed patients with anxiety or agitation. More recently, several SSRI antidepressants such as paroxetine and escitalopram have been evaluated in the treatment of various anxiety disorders. SSRIs are now the treatment of choice in chronic anxiety disorders.[22,23]

MAOIs have been used for many years to treat phobic states, but the well-known range of unwanted effects, including hypotension, limb oedema, and dietary and drug interactions, preclude their routine use.

β-adrenoceptor antagonists

Beta-adrenoceptor antagonists may help patients with anxiety, but usually only those complaining of somatic symptoms. They are still favoured by some primary care doctors.

Antihistamines

The older compounds penetrate the brain readily and are quite sedative. Hydroxyzine has been evaluated in at least two placebo-controlled trials in doses of 50 mg/day: it proved to be significantly better than placebo. The advantages are that paradoxical reactions are rare, cognitive function including memory is largely unaffected, and rebound and withdrawal seem rare.

Clinical effects of hypnotics

Insomnia is a common symptom,[24] especially in the elderly.[25] Nonetheless, many complaints of insomnia are unfounded as the patient has unreal expectations concerning sleep. Elderly people fail to appreciate that it is normal to sleep less and less deeply as they age. Napping during the day also decreases the need for sleep at night. Some people can manage on 5 to 6 h a night indefinitely, and yet worry that this is insufficient. Explanation and reassurance relieve their worries.

In many patients complaining of more severe insomnia, the cause is a physical complaint such as pain, breathlessness, or pruritus. The treatment is for that of the primary complaint. In many other cases, the insomnia is either a symptom of psychiatric distress, anxiety, or depression, or it is iatrogenic, caused by the very drugs prescribed to relieve the insomnia. In the first instance, treatment is directed towards the primary condition; in the second, a careful regimen of drug withdrawal, or substitution and subsequent withdrawal, should be instigated, as discussed earlier for anxiolytic medication. Some drugs, of which caffeine is the most common, induce insomnia. Alcohol may also disrupt sleep, particularly during the latter half of the night.

Despite this, a substantial number of patients cannot be placed into these categories and yet they persistently complain of insomnia (primary insomnia). Careful evaluation of the issues may yet reveal some relationship to stresses, both transient and persistent. It can be established that these patients are responding to unusual or protracted pressures of life: a man worries over possible redundancy, his wife is concerned about their delinquent son, their daughter is lovelorn, and grandmother is anxious over her increasing frailness. Giving drugs may set in train a long-term process culminating in drug-related insomnia without solving the basic problems.

Short-term symptomatic relief is acceptable when the stress is undoubtedly severe but transient. Even so, the hypnotic agent must be chosen carefully. The elimination half-life is the most important consideration. Those with half-lives over 12 h, such as nitrazepam, are only appropriate where an anxiolytic effect is required during the day as well as sleep induction at night. Even here, diazepam 5 to 15 mg, one dose at night, may be preferred. Temazepam with its shorter half-life will encourage sleep onset without leaving the patient with too many residual sedative effects the next day. Unfortunately, it has been extensively abused.

The management of chronic insomnia is much more problematic.[26] The newer compounds zopiclone and zolpidem are also short-acting agents and can help assure a good night's

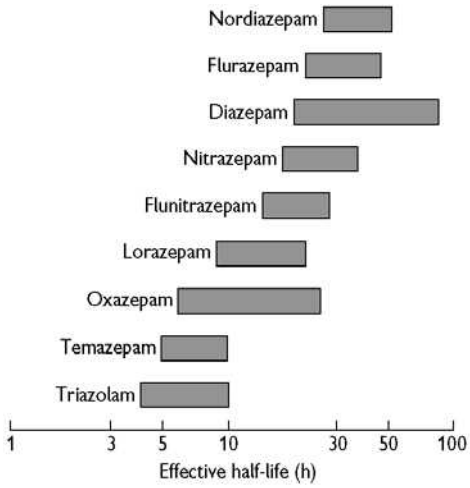


Fig. 6.2.2.2 Table of half-lives.

sleep without much risk of residual sedative effects the next day.[27] This is dependent on the dosage being kept modest, especially in the elderly. Eszopiclone is licensed in the United States for long-term use in chronic insomnia and is already used extensively there.

Zolpidem and zaleplon can be used in a different strategic ways from other longer-acting drugs. Hypnotics are traditionally taken every night before going to bed to induce or maintain sleep. However, the severity of insomnia usually varies from night to night. Consequently, regular usage may be partly, or even largely, unnecessary, and increases the risk of habituation and dependence. Very-short-acting compounds are unlikely to leave residual effects the next day, even taken up to 5 h or so before the expected time of wakening. Consequently, the insomniac can refrain from regular hypnotic usage but, instead, wait up to an hour or so after going to bed to see if natural sleep supervenes before resorting to medication. This changes the regular prophylactic use of hypnotics to 'as needed', and lessens the risk of habituation and dependence. Furthermore, the patient is gratified to feel that he or she has control of the medication instead of vice versa.

Residual effects

Residual effects can be a problem especially when long-acting drugs are used repeatedly. Dosage is important here, since residual effects increase in both magnitude and duration as the dose is increased. It should be remembered that hypnotics are the only class of drugs in which the main therapeutic effect (drowsiness) is identical with the main unwanted effect; the two are merely separated by 8 h in time. A short-acting hypnotic compound will be devoid of residual effects the next day, but the patient may wake early. After taking a longer-acting compound, sleep may be prolonged but hangover effects pronounced.

Idiosyncratic effects

Adverse effects with triazolam alerted prescribers and regulators to possible major adverse effects of short-acting benzodiazepines. The adverse reactions in question include daytime anxiety, amnesic effects, and episodes, and morbid affects such as depression and hostility. In summary, the evidence suggests that these are class effects common to the benzodiazepines, although more likely to occur the shorter the duration of action of the drug and the higher the dose. Alcohol is also capable of producing these effects.

Rebound

Discontinuation of many hypnotics is often followed by worsening of sleep compared with pretreatment levels. In practical terms, insomniac patients find that their sleep is disturbed for a night or two after abrupt discontinuation of what appeared to be effective medication. Some of this rebound is subjective as patients taking sleeping pills tend to overestimate their sleeping time (compared with sleep laboratory recordings); on withdrawal, they underestimate their sleep. The intensity of rebound insomnia is strongly related to dose but less clearly to the duration of use, and marked individual differences exist. The risk of rebound is greater with short-half-life compared with the long-half-life compounds. Tapering off medication lessens the likelihood of rebound. However, despite clinical impressions that rebound insomnia might lead to the resumption of medication, there is little evidence for this.[28]

Dependence

Dependence may supervene on the longer-term use of hypnotics; giving a long-acting benzodiazepine drug only once in 24 h does not protect against such an eventuality. The management of the withdrawal syndrome that may occur is largely the same as with the anxiolytic benzodiazepines.

Abuse

A growing problem with these drugs is abuse—non-medical use, on a regular or sporadic basis, often in a polydrug context. Worldwide, flunitrazepam is the main problem and can be taken orally, by injection, or by sniffing. In the United Kingdom, temazepam is widely abused by injection. The injected drug has a marked sedative and/or disinhibiting effect, resulting in chaotic behaviour, carelessness, and an enhanced risk of the transmission of communicable diseases such as HIV infection and hepatitis.

Other hypnotics

Gaboxadol is a GABA agonist. Preliminary data suggest useful hypnotic properties.

Melatonin preparations

A series of compounds are being developed based on melatonin. This hormone is important in the regulation of sleep and is secreted at night. Some elderly insomniacs seem to be deficient in melatonin. Preparations include Circadin®, licensed in the UK for short-term use in insomnia in over 55s. Ramelteon is a melatonin$_1$ and melatonin$_2$ agonist. It is licensed in the United States for insomnia, and is effective in inducing sleep: it has a favourable safety profile.[29]

Conclusions

In many countries the drug treatment of both anxiety and insomnia still largely revolves around the use of the benzodiazepines. Nevertheless, controversy and disagreement still rage about the risk–benefit ratio of compounds in this area. Short-term use in both indications is well established, with a favourable database as a

rationale for this approach. However, long-term use is still only researched in a limited way. While both the efficacy and safety of long-term use remain unclear, acceptance of current guidelines limiting the use of benzodiazepines seems wise.

The advent of the SSRIs as anxiolytics has driven a wedge between the treatment methods for anxiety and insomnia. Anxiety can be treated just as effectively with an SSRI (and probably, pregabalin) as with a benzodiazepine, and more safely. The treatment of insomnia still relies on the benzodiazepines until the risk–benefit ratio of newer drugs such as the melatonin-related compounds becomes clear.

Nevertheless, in the author's opinion the most important outstanding issue is the relationship between drug and non-drug treatments.[30] The management of anxiety disorders and of insomnia is complex and is hampered by a dearth of information concerning the relative merits of various treatment modalities. Much research is also needed on the optimum strategies for combining all the therapies available to us, and on identifying predictors of response.

Developments in the neuropharmacology of insomnia hold out the promise of new compounds with novel and perhaps more effective modes of action.[31] With respect to anxiety disorders, a major shift of emphasis has followed the demonstration of the efficacy of the SSRIs.[32]

Further information

Taylor, D., Kerwin, R., and Paton, C. (2005). *The Maudsley prescribing guidelines* (8th edn). South London and Maudsley NHS Trust.

Nutt, D. and Ballenger, J. (eds.) (2003). *Anxiety disorders*. Blackwell Science, Oxford.

References

1. Allgulander, C. (1986). History and current status of sedative-hypnotic drug use and abuse. *Acta Psychiatrica Scandinavica*, **73**, 465–78.
2. Lader, M. (1994). Benzodiazepines. A risk–benefit profile. *CNS Drugs*, **1**, 377–87.
3. Curran, H.V. (1991). Benzodiazepines, memory and mood: a review. *Psychopharmacology*, **105**, 1–8.
4. Barker, M.J., Greenwood, K.M., Jackson, M., *et al.* (2004). Cognitive effects of long-term benzodiazepine use. *CNS Drugs*, **18**, 38–45.
5. Langtry, H.D. and Benfield, P. (1990). Zolpidem. A review of its pharmacodynamic and pharmacokinetic properties and therapeutic potential. *Drugs*, **40**, 291–313.
6. Sramek, J.J., Zarotsky, V., and Cutler, N.R. (2002). Generalised anxiety disorder. Treatment options. *Drugs*, **62**, 1635–45.
7. Taylor, C.B. (2006). Panic disorder. *British Medical Journal*, **332**, 951–5.
8. Hoehn–Saric, R. (1998). Generalised anxiety disorders. Guidelines for diagnosis and treatment. *CNS Drugs*, **9**, 85–98.
9. Otto, M.W., Pollack, M.H., and Sachs, G.S. (1993). Discontinuation of benzodiazepine treatment: efficacy of cognitive–behavioral therapy for patients with panic disorder. *The American Journal of Psychiatry*, **150**, 1485–90.
10. Dietch, J.T. and Jennings, R.K. (1988). Aggressive dyscontrol in patients treated with benzodiazepines. *The Journal of Clinical Psychiatry*, **49**, 184–7.
11. Serfaty, M. and Masterton, G. (1993). Fatal poisonings attributed to benzodiazepines in Britain during the 1980s. *The British Journal of Psychiatry*, **163**, 386–93.
12. Michelini, S., Cassano, G.B., Frare, F., *et al.* (1996). Long-term use of benzodiazepines: tolerance, dependence and clinical problems in anxiety and mood disorders. *Pharmacopsychiatry*, **29**, 127–34.
13. Petursson, H. and Lader, M.H. (1981). Withdrawal from long–term benzodiazepine treatment. *British Medical Journal*, **283**, 643–5.
14. Tyrer, P. (1991). The benzodiazepine post-withdrawal syndrome. *Stress Medicine*, **7**, 1–2.
15. Russell, J. and Lader, M. (eds.) (1993). *Guidelines for the prevention and treatment of benzodiazepine dependence*. Mental Health Foundation, London.
16. Voshaar, R.C., Couvee, J.E., van Balkom, A.J., *et al.* (2006). Strategies for discontinuing long-term benzodiazepine use: meta-analysis. *The British Journal of Psychiatry*, **189**, 213–20.
17. Potokar, J. and Nutt, D.J. (1994). Anxiolytic potential of benzodiazepine receptor partial agonists. *CNS Drugs*, **1**, 305–15.
18. Fulton, B. and Brogden, R.N. (1997). Buspirone. An updated review of its clinical pharmacology and therapeutic applications. *CNS Drugs*, **7**, 68–88.
19. Kavoussi, R. (2006). Pregabalin: from molecule to medicine. *European Neuropsychopharmacology*, **16**, S128–33.
20. Frampton, J.E. and Foster, R.H. (2006). Pregabalin in the treatment of generalised anxiety disorder. *CNS Drugs*, **20**, 685–93.
21. Van Ameringen, M., Mancini, C., Pipe B., *et al.* (2004). Antiepileptic drugs in the treatment of anxiety disorders. Role in therapy. *Drugs*, **64**, 2199–220.
22. Stein, D.J. (2006). Evidence-based treatment of anxiety disorders. *International Journal of Psychiatry in Clinical Practice*, **10**(Suppl. 1), 16–21.
23. Baldwin, D.S., Anderson, I.M., Nutt, D.J., *et al.* (2005). Evidence-based guidelines for the pharmacological treatment of anxiety disorders: recommendations from the British Association for Psychopharmacology. *Journal of Psychopharmacology*, **19**, 567–96.
24. Üstün, T.B., Privett, M., Lecrubier, Y., *et al.* (1996). Form, frequency and burden of sleep problems in general health care: a report from the WHO collaborative study on psychological problems in general health care. *European Psychiatry*, **11**(Suppl. 1), 5S–10S.
25. Kamel, N. and Gammack, J.K. (2006). Insomnia in the elderly: cause, approach, and treatment. *The American Journal of Medicine*, **119**, 463–9.
26. Kupfer, D.J. and Reynolds, C.F. (1997). Management of insomnia. *The New England Journal of Medicine*, **336**, 341–6.
27. Nowell, P.D., Mazumdar, S., Buysse, D.J., *et al.* (1997). Benzodiazepines and zolpidem for chronic insomnia. A meta-analysis of treatment efficacy. *The Journal of the American Medical Association*, **278**, 2170–7.
28. Roehrs, T., Merlotti, L., Zorick, F., *et al.* (1992). Rebound insomnia and hypnotic self administration. *Psychopharmacology*, **107**, 480–4.
29. Johnson, M.W., Suess, P.E., and Griffiths, R.R. (2006). Ramelteon: a novel hypnotic lacking abuse liability and sedative adverse effects. *Archives of General Psychiatry*, **63**, 1149–57.
30. Fineberg, N. and Drummond, L.M. (1995). Anxiety disorders. Drug treatment or behavioural cognitive psychotherapy? *CNS Drugs*, **3**, 448–66.
31. Szabadi, E. (2006). Drugs for sleep disorders: mechanisms and therapeutic prospects. *British Journal of Clinical Pharmacology*, **61**, 761–6.
32. Tyrer, P. and Baldwin, D. (2006). Generalised anxiety disorder. *Lancet*, **368**, 2156–66.

6.2.3 Antidepressants

Zubin Bhagwagar and George R. Heninger

Introduction

Major depressive disorder is a serious, recurrent illness which levies a crippling toll on individuals, families, and society in general. The importance of depression as a major public health problem is emphasized by findings from the World Health Organization Global Burden of Disease survey in showing that in 1990 it was the fourth largest cause of burden of disease (i.e. years of life lost due either to premature mortality or to years lived with a disability). It has been estimated that by the year 2020 it is expected to be the second largest cause of burden of disease.[1] Depression is underdiagnosed and frequently under-treated, and depressed individuals have a much higher risk for suicide. The primary treatment for depression involves the use of antidepressant drugs, and it is therefore important that clinicians become familiar with and adept in utilizing this important group of compounds. Although primarily used for the treatment of depression, drugs within this category also have a number of other important uses. A thorough understanding of the pharmacology of antidepressants will aid the clinician in the selective use of these drugs for patients with depression as well as patients with a number of other disorders.

A brief history of antidepressant discovery and theories of action

Table 6.2.3.1 gives a brief chronology of antidepressant drug discoveries and theories of drug action. It is a comment on our understanding of the illness that some of the major advances in the pharmacotherapy of depression have been serendipitous. Prior to 1954, except for the use of electroconvulsive therapy, there were few effective drug treatments for depression. In 1954, the antidepressant era was initiated with the observation that some patients with tuberculosis displayed mood elevations following treatment with the antituberculosis agent iproniazid.(2) Following this initial serendipitous observation, the antidepressant effect of iproniazid was confirmed(3) and its action of inhibiting monoamine oxidase was reported. Iproniazid had significant toxicity and other monoamine oxidase inhibitors (MAOIs) were subsequently introduced. Independent from the work on MAOIs, imipramine, which has a chemical structure similar to the phenothiazines, was assessed as an agent to treat agitation in psychotic patients where it was found to be ineffective. However, it was noticed (again serendipitously) that imipramine produced an improvement of mood in the subset of patients who had symptoms of depression. Kuhn then reported in 1958 that imipramine was an effective antidepressant.(4)

One of the earliest theories of antidepressant drug action was that the antidepressant effect was produced by an increase of serotonin (5-hydroxytryptamine (**5-HT**)) in brain. This was supported by an initial study showing that an MAOI plus tryptophan, the precursor of 5-HT, was a more effective antidepressant treatment than an MAOI alone.[5] Subsequently, the discovery that imipramine and desipramine had effects in inhibiting the reuptake of noradrenaline (norepinephrine) and adrenaline (epinephrine) into the synapse led to the catecholamine theory of depression, which proposed that antidepressant treatments act by increasing the level of catecholamines at brain synapses.[6,7] Ten years later, it was reported that in laboratory animals most antidepressant treatments lead to downregulation of β-adrenergic receptors. This supported the proposal that antidepressants act by reducing β-adrenergic receptor sensitivity.[8] However, the reduction in β-adrenergic receptor sensitivity occurred within hours and antidepressant effect requires 1 to 3 weeks and futher not all effective antidepressant treatments produce reductions in β-adrenergic receptor sensitivity.

In the 1970s and 1980s, a large number of studies on antidepressants were conducted in laboratory animals which demonstrated that they produced a number of changes in monoamine receptor sensitivity.[9] In the late 1980s, a number of neurophysiological

Table 6.2.3.1 History of discovery of antidepressants and pharmacological theories of antidepressant drug action

Year	Discovery or theory	Reference
1954	Discovery that MAOIs have antidepressant effects	2, 3
1958	Discovery that the tricyclic drug imipramine is an effective antidepressant	4
1963	*Serotonin theory of depression:* MAOIs act by increasing serotonin and tryptamine in brain	5
1965	*Catecholamine theory of depression:* ADTs act by increasing cathecolamines in brain	6, 7
1975	β-*Adrenergic receptor theory of depression:* ADTs act by altering the sensitivity of several monoamine receptor subtypes in brain	8
1981	*Monoamine receptor sensitivity theory of depression:* ADTs act by altering the sensitivity of several monoamine receptor subtypes in brain	9
1987	*Serotonergic augmentation theory of depression:* ADTs act by decreasing sensitivity of presynaptic serotonergic autoreceptors and increasing sensitivity of serotonergic postsynaptic receptors to increase overall efficacy in serotonergic transmission	10
1996	*A molecular and cellular theory of depression:* ADTs act by producing a sustained activation of the CAMP system which increases brain levels of neurotrophic factors that reverse the effects of stres in certain brain areas	14
1998	Discovery that a substance P antagonist that does not interact with monoamine systems is as effective an antidepressant as an SSRI (paroxetine)	16
2000/2006	Demonstration of antidepressant properties of ketamine implicating the glutamatergic system in the pathophysiology of depression	19, 20

ADT, antidepressant treatment.

studies provided evidence that the delay in onset of antidepressant effects could be accounted for by a slow decrease in sensitivity at presynaptic serotonergic autoreceptors which has the overall result of increasing serotonergic function after days and weeks of treatment.[10] An elaboration on the receptor sensitivity theory was the discovery that most antidepressants produce alterations in the sensitivity of a specific glycine-sensitive site on the *N*-methyl-d-aspartate (NMDA) receptor.[11] A subsequent study showed that an NMDA antagonist may have antidepressant actions[12]and this line of thought has borne fruit recently in a possible novel mechanism of action for antidepressant treatment. An additional receptor sensitivity change thought to be important in the mechanism of action of antidepressants involved changes in the sensitivity of receptors for glucocorticoids. It was found that antidepressants produce an overall improvement of inhibitory feedback on the hypothalamic-pituitary-adrenal axis[13] and that specific corticotrophin releasing hormone (CRH) antagonists have antidepressant properties.[14]

A more recent theory of antidepressant drug action involves findings that antidepressant treatments affect intracellular pathways and neurotrophins. It was found that many antidepressants, in spite of β-adrenergic receptor downregulation, continue to produce sustained activation of the cAMP system and that this is related to increases of neurotrophic factors in brain.[15] Neurotrophins reverse the effects of stress in some brain areas and this raise the possibility that antidepressants act by increasing neurotrophins which reverse the effects of stress in important brain areas of depressed patients.

Throughout the 1980s, and 1990s a number of compounds that do not fit the standard monoamine theories of depression have been found to be effective clinical antidepressants. One of these drugs, tianeptine, actually increases the uptake of 5-HT into nerve endings, an effect that is opposite to the standard selective serotonin reuptake inhibitors (**SSRI**s).[16] Similarly, while there was intense interest in a report of possible antidepressant efficacy of a substance P receptor antagonist,[17] which does not interact with monoamine systems, clinical trials for this specific compound were disappointing.

Although no single mechanism has been discovered that will account for the antidepressant effects of all effective antidepressant treatments, it is clear that initial effects on monoamine metabolism with subsequent effects of intracellular pathways is important. While clinical wisdom and data suggest that there is a lag of 7–21 days to antidepressant action, recent reports question this notion of delayed onset of efficacy.[18] Recently ketamine, an NMDA antagonist has been shown to have an onset of action much faster than that traditionally seen with conventional antidepressants[19,20] suggesting that the pursuit of novel mechanisms may indeed result in advances in the pharmacotherapy of depression. Using preclinical models, it has been suggested that both nonselective NMDA antagonists as well as NR2B selective antagonists exert their antidepressant effects by regulating the functional interplay between AMPA and NMDA throughput.[21]

Pharmacology and types of compounds available

Antidepressant drugs fall into a wide variety of chemical classes and they have a wide range of neuropharmacological effects. They are grouped in Tables 6.2.3.2, 6.2.3.3, and 6.2.3.4 based on the presumed primary action that leads to an antidepressant effect. Table 6.2.3.2 lists the drugs that inhibit the uptake of the monoamines noradrenaline, 5-HT, and dopamine into nerve endings which in turn is thought to increase the function of the respective monoamine systems in brain. Table 6.2.3.3 lists the drugs that inhibit monoamine oxidase and thereby increase the concentration of many amines in brain. Table 6.2.3.4 lists the drugs with other primary actions that do not primarily involve inhibition of monoamine uptake or monoamine oxidase inhibition.

In Table 6.2.3.2, the first 12 compounds are inhibitors of noradrenaline uptake with a variable potency of inhibiting 5-HT uptake. The drugs with secondary amine structures, desipramine, nortriptyline, protriptyline, amoxapine, and maprotiline are predominantly noradrenaline uptake inhibitors with little effect on 5-HT uptake.[22] It can be seen in Table 6.2.3.2 that clomipramine, in addition to inhibiting noradrenaline uptake, is also a strong 5-HT uptake inhibitor. There are currently three selective serotonin and noradrenaline reuptake inhibitor (SNRI) drugs available; milnacipran (not licensed in the US), venlafaxine and duloxetine. Venlafaxine, inhibits both 5-HT and noradrenaline and 5-HT reuptake,[23] as do milnacipran and duloxetine though in varying proportions. While affinities vary depending on the system studied, milnacipran blocks 5-HT and norepinephrine reuptake with relatively equal affinity, while duloxetine has been suggested to have a slightly greater selectivity for 5-HT and venlafaxine a much greater selectivity for 5-HT.[24] Reboxetine (not licensed in the US) is a highly selective and potent inhibitor of noradrenaline reuptake.[25] It has only a weak effect on the 5-HT reuptake and does not affect the uptake of dopamine.

A key issue in the pharmacology of all antidepressant drugs is the relative specificity of their action. Drugs with a tertiary amine structure tend to produce more antagonism of α_1-adrenergic receptors which can produce hypotension, histamine receptors which can produce sedation, and muscarinic cholinergic receptors which can produce blurred vision, dry mouth, and urinary retention. This leads to more side-effects for these compounds than the drugs with a secondary amine structure. Venlafaxine has relatively less effect on these receptors and thus fewer side-effects[23] (see Table 6.2.3.6).

SSRIs are probably the most widely prescribed antidepressants and represent a class of drugs that selectively inhibit 5-HT reuptake from the synapse. Unlike the tricyclics, they each have different chemical structures. The drugs listed in Table 6.2.3.2 have a relatively specific effect in inhibiting 5-HT uptake,[22] and because of their relatively specific effect on this monoamine system and the lack of antagonism of many other receptors, they have been found to have fewer side-effects. Escitalopram was introduced following the discovery that all of the inhibitory activity of citalopram on 5-HT reuptake resides in the *S*-(+)-enantiomer (*S*-citalopram),[26] with *S*-citalopram being 167 times more potent than *R*-citalopram at inhibiting 5-HT reuptake into rat brain synaptosomes.

MAOIs are listed in Table 6.2.3.3. Two isozymes, monoamine oxidases A and B, are present in many discrete cell populations within the central nervous system, and glial cells also express monoamine oxidases A and B. The main substrates for monoamine oxidase A include adrenaline, noradrenaline, and 5-HT. The breakdown of dopamine in striatal regions of the brain is preferentially by monoamine oxidase B, but it can also be broken down by monoamine oxidase A. Since monoamine oxidase is located on the

Table 6.2.3.2 Pharmacological actions of antidepressants: drugs that inhibit monoamine reuptake at the synapse

Drug	Chemical class	Relative reuptake inhibition		
		Noradrenaline	5-Hydroxytryptamine	Dopamine
Imipramine	Tricyclic	++	+	0
Desipramine[a]	Tricyclic	++++	0	0
Amitriptyline[a]	Tricyclic	++	+	0
Nortriptyline[a]	Tricyclic	+++	0/+	0
Trimipramine	Tricyclic	+	0	0
Clomipramine	Tricyclic	+	+++	0
Protriptyline[a]	Tricyclic	++++	0	0
Doxepin	Tricyclic	++	0/+	0
Amoxapine[a]	Tricyclic	+++	0	+
Maprotiline[a]	Tetracyclic	+++	0	0
Venlafaxine	Bicyclic	+	++	0/+
Milnacipran	SNRI	+++	+++	0/+
Duloxetine	SNRI	+++	++	0/+
Reboxetine	NARI	++++	0	+
Fluoxetine	SSRI	0	+++	0
Sertraline	SSRI	0	++++	+
Fluvoxamine	SSRI	0	+++	0
Paroxetine	SSRI	+	++++	0
Citalopram	SSRI	0	++++	0
Escitalopram	SSRI	0	++++	0

0, None; 0/+, minimal; +, low; ++, moderate; +++, High; ++++, very high.

[a]Secondary amine.

outside of the plasma membrane of the mitochondria in neurones, it is not able to eliminate amines that are stored inside vesicles. MAOI produces an increase in monoamines in the cytoplasm. It is thought that the increase in monoamine content is the primary mechanism of action of MAOIs, and other secondary changes including β-adrenergic receptor downregulation and other receptor changes are secondary to the increased amine levels.[27]

Four of the six drugs listed in Table 6.2.3.3 are irreversible inhibitors. The two reversible inhibitors are essentially inert substrate analogues, and there is usually a correlation between their plasma concentration and the reversible inhibition of monoamine oxidase A. Since isocarboxazid, phenelzine, and tranylcypromine are irreversible inhibitors of monoamine oxidases A and B, there can be serious side-effects when foods that are high in tyramine or other amines are ingested. In addition, these three drugs have strong interactions with other drugs that alter monoamine

Table 6.2.3.3 Pharmacological actions of antidepressants: drugs that inhibit monoamine oxidase

Drug	Chemical class	MAO A	MAO B	Reversible
Isocarboxazid	Hydrazine	Yes	Yes	No
Phenelzine	Hydrazine	Yes	Yes	No
Tranylcypromine	Amphetamine	Yes	Yes	No
Moclobemide	Morpholine	Yes	No	Yes
Brofaromine	Piperidine	Yes	No	Yes
Selegiline	Phenethylamine	No	Yes[a]	No

MAO, Monoamine oxidase.

[a]Selective at lower doses; becomes non-selective at higher doses.

Table 6.2.3.4 Pharmacological actions of antidepressants: drugs that do not act by strong inhibition of monoamine uptake or inhibition of monoamine oxidase

Drug	Chemical class	Possible pharmacological action
Trazodone	Triazolopyridine	Mixed 5-HT agonist/antagonist
Nefazodone	Phenylpiperazine	Mixed 5-HT agonist/antagonist, weak monoamine uptake inhibitor
Bupropion	Unicyclic amino ketone	Weak noradrenaline and dopamine uptake inhibitor
Mianserin	Tetracyclic	Antagonist α_2-adrenergic auto- and heteroreceptors, increased 5-HT and noradrenaline release
Mirtazapine	Tetracyclic	Antagonist α_2-adrenergic auto and heteroreceptors, increased 5-HT and noradrenaline release

metabolism and therefore their use as antidepressants is much more limited than the tricyclics, SSRIs, or other antidepressant compounds. Tranylcypromine, which has a structure similar to amfetamine in addition to being an MAOI, is also thought to have a stimulant-type action of rapid onset. With the reversible MAOIs moclobemide and brofaromine, the recovery of monoamine oxidase back to normal levels after the drug is stopped is much shorter than with the irreversible MAOIs. These drugs increase concentrations of 5-HT, noradrenaline, and adrenaline that are short and parallel the time course of the monoamine oxidase A inhibition. These two drugs are more easily displaced by the pressor amines such as tyramine, and therefore, are thought to be safer than the irreversible inhibitors.

Selegiline, which has recently become available as a transdermal patch,[28] is selective at lower doses for monoamine oxidase B but at higher doses it becomes non-selective.[29] It has been primarily used for the treatment of Parkinson's disease and the doses for treating depression need to be much higher (note: selegiline is not licensed in the UK for depression). Since monoamine oxidase B is not involved in the intestinal tyramine interaction, selegiline interactions with ingested monoamines have been minimal.

In addition to inhibiting monoamine oxidase, these compounds have other effects on monoamine systems that can produce side-effects. However, the major concerns are the interactions with dietary amines and other drugs that influence amine function. The combination of dietary interactions and slow recovery of monoamine oxidase following with the irreversible inhibitors makes these drugs one of the more difficult treatments to administer. They are generally reserved for patients not otherwise responding to the other less toxic antidepressants.

In Table 6.2.3.4, compounds that are effective antidepressants but do not inhibit monoamine oxidase or have strong monoamine uptake inhibition are listed. Trazodone has shown receptor antagonist activity at several 5-HT receptor subtypes although its active metabolite *m*-chlorophenylpiperazine (mCPP) is a potent direct serotonin agonist. It is a weak but relatively selective inhibitor of 5-HT reuptake, is an antagonist at 5-HT1A and 5-HT2 receptors in addition to its active metabolite mCPP being a potent 5-HT agonist.[30] This leads to trazodone being classified as a mixed 5-HT agonist/antagonist. It also has relatively weak 5-HT uptake inhibiting properties but with no effect on noradrenaline or dopamine uptake. Trazodone is virtually devoid of anticholingeric activity and therefore it has few side-effects in this area. However, it does produce considerable sedation and hypotension secondary to antagonism of α_1-adrenergic receptors and histamine receptors.

Nefazodone is an analogue of trazodone that was developed to overcome the orthostatic hypotension and sedation caused by the latter. Like trazodone it is a 5-HT receptor antagonist with weak monoamine uptake inhibition activity.[31] It has less affinity for the α-adrenergic receptors and is inactive on many other receptors. It too is metabolized to *m*-chlorophenylpiperazine which is an active serotonergic agonist. Although the initial effects of nefazodone involve alterations of 5-HT neurotransmission, these effects are complex and depend on the biological test used.

Bupropion resulted from focussed research to find antidepressant compounds that would have fewer side-effects than traditional tricyclics (note: buproprion is not licensed in the UK for depression). Bupropion is a mild inhibitor of noradrenaline uptake, has some effects on inhibiting dopamine uptake but has no effect on 5-HT uptake.[32] These effects are not associated with β-adrenergic receptor downregulation as is seen with many other antidepressants. One of the active metabolites is hydroxybupropion which also has an antidepressant profile in laboratory animals. It is of interest that bupropion is one of the few drugs that reduce REM latency since most other treatments increase it. Although the specific mechanisms of bupropions antidepressant effects are not known, its unique profile has led to its use in the treatment of bipolar disorder[33] as well as its use in the treatment of smoking cessation.[34]

Mianserin and mirtazapine both have potent effects on antagonizing α_2-adrenergic auto- and heteroreceptors.[35] They also antagonize other 5-HT receptors but have minimal effects on monoamine uptake or monoamine oxidase activity. Since α_2 receptors inhibit noradrenaline release, their antagonism leads to an increase in noradrenaline release in many brain areas. In addition, antagonism of α_2-adrenergic heteroreceptors located on serotonergic neurones results in an enhanced 5-HT release. With mirtazapine, since 5-HT_2 and 5-HT_3 receptors are blocked, this could result in selective enhancement of 5-HT_1-receptor-mediated neurotransmission. These drugs have low affinity for muscarinic, cholinergic, and dopamine receptors and this is related to a reduced side-effect profile. The combination of increased noradrenaline release and increased 5-HT release resulting from the α_2-antagonism on auto- and heteroreceptors is hypothesized to be the central mechanism of action.

Pharmacokinetics

Data on the pharmacokinetics of antidepressants are listed in Table 6.2.3.5. The tricyclic antidepressants are by and large well absorbed although time to peak plasma concentration can vary from 1 to 12 h depending on the drug and the individual. In general, these drugs are metabolized in the liver to a variety of metabolites, some of which are active. For instance, desipramine is a metabolite of imipramine and nortriptyline is a metabolite of amitriptyline. Most of these compounds have a long half-life (close to 24 h) that will allow for once-daily dosing. All the compounds are highly bound to plasma protein except for venlafaxine and milnacipran. Although many of the compounds have active metabolites, the exact percentage of each metabolite in patients and in their clinical effects is still largely unknown.

There is considerable individual variation in the metabolism of tricyclics, and a large component of this may be genetic. Up to 7 to 9 per cent of the Caucasian population have been classified as slow metabolizers (slow hydroxylators) which can be measured by the rate of hydroxylation of debrisoquin. The slow hydoxylation has been determined to be caused by a polymorphism in a cytochrome P-450 macrosomal enzyme (CYP2D6). It is of interest that many SSRIs are inhibitors of P-450 isoenzymes which can considerably influence the metabolism of tricyclic antidepressants.[36] In general, the increased renal clearance in children and a decreased renal clearance with age need to be taken into account with dosing.

The SSRIs are rapidly absorbed, although there is variability within the drug half-lives. The metabolism into active metabolites can vary the pharmacodynamic effects considerably. For example, fluoxetine is metabolized to norfluoxetine which has similar activity on 5-HT reuptake as fluoxetine. The elimination half-life of norfluoxetine is longer (4–16 days) than that of fluoxetine (4–6 days). The desmethyl metabolite of sertraline although not nearly as

potent as the parent compound, also has a much longer half life. The desmethyl metabolite of citalopram or escitalopram, although a potent noradrenaline uptake inhibitor, is much lower in concentration than citalopram and it weakly crosses the blood-brain barrier. Fluvoxamine, paroxetine, duloxetine or milnacipran do not have any active metabolites. The relatively long half-lives of some of the SSRIs, particularly fluoxetine, require longer drug-free periods before switching to other classes of compounds especially before starting an MAOI.

The MAOIs are all rapidly absorbed. For the irreversible MAOIs, the elimination half-life and protein binding patterns are not as relevant because of the irreversible effects on monoamine oxidase. The reversible MAOIs have shorter half-lives and require multiple daily dosing.[29] With the irreversible MAOIs, once the drug is stopped, there needs to be time for new synthesis of monoamine oxidase. This requires a minimum of 5 to 7 days and the safest recommendation is to wait 2 weeks before starting other drugs that may interact with the MAOIs.

The five other antidepressants listed in Table 6.2.3.5 are rapidly absorbed but there is some variation in their elimination half-life. In general, the half-lives are short enough that multiple daily dosing is required. They are generally bound to plasma protein at a high level. The metabolites of trazodone and nefazodone have mixed effects on 5-HT receptors which results in a complex overall effect. Trazodone and nefazodone undergo extensive hepatic metabolism and one major metabolite is *m*-chlorophenylpiperazine which stimulates 5-HT receptors. Many metabolites have biological activity with half-lives different to the parent compounds.

Bupropion is metabolized in the liver and its metabolites can be at higher concentration than the parent compound. The relationship between plasma bupropion and clinical response has been poor.

Table 6.2.3.5 Pharmacokinetics of antidepressants

Drug	Absorption time to peak plasma concentration (h)	Elimination half-life (h)	Percentage plasma protein binding	Important metabolite
Monoamine reuptake inhibitors				
Imipramine	1.5–3	11–25	92	Desipramine
Desipramine	3–6	11–31	90	2-OH-desipramine
Amitriptyline	1–5	10–26	94	Nortriptyline
Nortrityline	3–12	18–44	92	10-OH-nortriptylene
Trimipramine	3	9–11	95	None
Clomipramine	2–6	21–31	97	Desmethylclomipramine
Protriptyline	6–12	67–89	93	None
Doxepin	1–4	11–23	80	Desmethyldoxepin
Amoxapine	1–2	8–30	90	8-OH-amoxapine
Maprotiline	4–12	28–58	88	Desmethylmaprotiline
Venlafaxine	2	5	30	O-desmethylvenlafaxine
Milnacipran	0.5–4	8	13	None
Duloxetine	6–10	8–17	95	None
Reboxetine	1.5–2.4	12–14	97	NA
SSRIs				
Fluoxetine	4–8	24–120	94	Norfluoxetine
Sertraline	6–8	27	99	n-Desmethylsertraline
Fluvoxamine	2–8	15–26	77	None
Paroxetine	5–7	24–31	95	None
Citalopram	1–6	33	80	NA (monodesmethylcitalopram)
Escitalopram	3–6	22–32	56	S-desmethylcitalopram
MAOIs				
Isocarboxazid	3–5	NA	NA	NA
Phenelzine	2–4	NA	NA	NA
Tranylcypromine	1.5–3	1.5–3.5	NA	NA
Moclobemide	1–1.5	1.4	NA	Numerous
Brofaromine	1–2	12–15	NA	n-Desmethylbrofaromine
Selegiline	1–3	2–10	NA	n-Desmethylselegiline
Other antidepressants				
Trazodone	1–2	6–11	92	m-Chlor ophenylpiperazine
Nefazodone	1	2–4	99	m-Chlorophenylpiperazine
Bupropion	3	10–21	85	Bupropion threoamino alcohol
Mianserin	2–3	15–22	NA	NA
Mirtazapine	2–3	20–40	85	None

NA, data not available

Side-effects

The history of new drugs becoming available for the treatment of depression reflects the efforts by the pharmaceutical industry to find compounds with reduced side-effects. This is in particularly important in the treatment of patients with medical illness because some of the side-effects can have considerable negative medical consequences. In Table 6.2.3.6, the propensity of the different drugs to produce some of the side-effects caused by antidepressants can be compared. The drugs that have high affinity for the α_1-adrenergic receptors can produce hypotension. Antagonism of histamine receptors has been associated with sedation and there is a long list of anticholinergic effects associated with antagonism of muscarinic cholinergic receptors.

For the tricyclics, it can be seen that drugs with a tertiary amine structure produce increased sedation. There is also an increase in the frequency of side-effects associated with antagonism of muscarinic cholinergic receptors such as dry mouth, constipation, blurred vision, urinary retention, dizziness, tachycardia, memory impairment, and at high and toxic doses, delirium. There is also an increased tendency for these same compounds to produce hypotension and to have unwanted cardiac effects that can lead to serious complications. In addition, the tertiary amines have a tendency to produce more weight gain than the secondary amines. The adverse effects have a particular impact on the tolerance of the patients to taking the medication. Most importantly the anticholinergic and cardiac effects can produce difficult complications in the elderly even leading to delirium when too high a dose is given. Amoxapine can cause extrapyramidal symptoms which are thought to be secondary to blocking dopamine receptors.[37] The most common adverse effects in patients taking reboxetine during clinical trials were insomnia, sweating, constipation, dry mouth, and urinary hesitancy compared with placebo, and the rates of nausea, diarrhea, and somnolence were lower compared with fluoxetine.[38] Nausea, dry mouth, dizziness, headache, somnolence, constipation, and fatigue were reported most frequently with duloxetine.[39]

The propensity to produce orthostatic hypotension is also a serious side-effect, particularly in the elderly. With the increased risk of falls and subsequent fractures in the elderly, this can be a serious health risk. A number of methods such as teaching patients to rise slowly from a supine position, tilting the bed upward, and maintenance of fluid uptake could help prevent this. However, other equally effective newer antidepressants produce much less of many of these side-effects, and they can be more safely used in the elderly.

Many of the drugs that are monoamine uptake inhibitors can cause cardiac conduction delays which may even lead to heart block in patients with pre-existing conditions. Severe overdose of these compounds can produce major and life-threatening cardiac arrhythmias. The secondary amines are generally thought to produce less cardiac effects than the tertiary amines. One of the characteristics of the SSRIs which has led to their widespread use is their low rate of side-effects. The pharmacological specificity of these compounds which bind to the 5-HT transporter, while not binding to the other neurotransmitter receptor types, results in their producing a therapeutic effect without many of the unwanted side-effects. In placebo-controlled trials the incidence of early discontinuation of SSRIs because of adverse events is intermediate between patients treated with placebo and patients treated with tricyclic antidepressants. Some of the symptoms reported with these compounds include agitation, anxiety, headache, sleep disturbance, and tremor. One of the more troublesome side-effects is sexual dysfunction especially anorgasmia. Less frequently, there are changes in appetite with nausea, dry mouth, sweating, and weight change. In general, these effects are less than those observed with the non-SSRI monoamine uptake inhibitors. The interaction of SSRIs with MAOIs to produce the serotonin syndrome is discussed under toxic effects below. Fluoxetine and sertraline induced higher rates of sedation as dosages are increased but in contrast, paroxetine produces a dose-dependent increase in arousal.

In contrast with the fewer side-effects produced by SSRIs and the other antidepressants that are not monoamine inhibitors, the MAOIs tend to produce frequent and often much more serious side-effects. Frequent side-effects include dizziness, headache, insomnia, dry mouth, blurred vision, nausea, constipation, forgetfulness, difficulty with urination, and weakness. There is also sexual dysfunction, including anorgasmia, impotence, delayed ejaculation, and decreased desire. Insomnia has also been reported. The original MAOIs iproniazid and isocarboxazid had a higher frequency of impairing liver function, but this is less with the other drugs. Pyridoxine deficiency has been reported and should be considered in evaluating side-effects. The largest problem with the MAOIs is the interactions with foods and with other drugs. Food interaction is much less of a problem with the reversible MAOIs moclobemide and brofaromine.[37–39]

Trazodone and nefazodone lack the anticholeringic side-effects of many of the tricyclic drugs. This makes them useful compounds in many medical conditions where this effect would be problematic. Trazodone has an acute sedative effect which is useful in the treatment of agitation, anxiety, and insomnia. However, this can be a troublesome side-effect when the patient performs tasks that require full alertness. Trazodone appears to have more propensity to produce orthostatic hypotension than nefazodone, possibility related to the degree of α_1-adrenergic receptor antagonism. Both trazodone and nefazodone, because of their lack of anticholinergic effects, have a low probability of producing difficulties in patients with cardiac illness. There is a slight tendency for weight gain but not nearly as strong as for some of the other antidepressants.[40] A relatively rare but important side-effect with trazodone is priapism. The risk for this side-effect is greatest during the early phase of treatment and the reporting of abnormal erectile function, including inappropriate or prolonged erections, should prompt quick discontinuation of trazodone treatment. Sexual side-effects have also been reported in women.

Bupropion has a very different side-effect profile than the conventional tricyclic antidepressants. It has no anticholinergic effects, is not sedating, and instead of weight gain, it suppresses appetite in some patients. In comparison to the SSRIs and trazodone and nefazodone, it also does not cause sexual dysfunction. There is no orthostatic hypotension, and bupropion does not produce cardiac side-effects. The possible stimulation of dopaminergic systems by bupropion can be related to its activating effects. This may be useful in patients with retardation but may exacerbate patients with agitation and insomnia. Bupropion can make tics in attention-deficit hyperactivity disorder and Tourette's disorder worse.[41] Patients have been described with bupropion-related

Table 6.2.3.6 Side-effects of antidepressants

Drug	Sedation	Anticholinergic effects	Hypotension	Cardiac effects	Weight gain
Monoamine reuptake inhibitors					
Imipramine	+++	++	++	+++	++
Desipramine	+	+	+	++	+
Amitriptyline	+++	++++	+++	+++	+++
Nortriptyline	+	+	+	++	+
Trimipramine	+++	+++	++	+++	++
Clomipramine	++	+++	++	+++	+
Protriptyline	0/+	++	+	+++	+
Doxepin	+++	++	+++	++	++
Amoxapine	+	++	+	++	+
Maprotiline	++	++	++	++	+
Venlafaxine	0/+	0/+	0	+	0
Milnacipran	0/+	+/++	0/+	+	0
Duloxetine	+	++	0/+	0/+	+
Reboxetine	0	+/++	0/+	0/+	0
SSRIs					
Fluoxetine	0/+	0	0	0	0
Sertraline	0	0	0	0	0
Fluvoxamine	0	0	0	0	0
Paroxetine	+	+	0	0	0
Citalopram	+	+	0/+	0/+	0
Escitalopram	+	+	0/+	0/+	0/+
MAOIs					
Isocarboxazid	+	+	+++	0	+
Phenelzine	+	0	+++	0	++
Tranylcypromine	+	0	++	0	0/+
Moclobemide	0	0	0	0	0
Brofaromine	0	0/+	0/+	0	0
Selegiline	0	0	+	0	0
Other antidepressants					
Trazodone	+++	0	++	0/+	+
Nefazodone	+	0	+	0/+	0/+
Bupropion	0	0	0	+	0
Mianserin	+++	0/+	0/+	+	+
Mirtazapine	+++	+	0/+	+	+

0, None; 0/+, 0ccasional; +, law; ++, moderate; +++, veryhigh.

psychosis which includes hallucinations and delusions. Psychotropic drugs also modulate seizure threshold and this needs to be carefully evaluated.[42] For example, a serious side-effect of bupropion that is rare but clinically important is the propensity to induce seizures in doses over 450 mg/day. Thus, bupropion should not be used at a dose higher than this and careful evaluation of history of seizures and other medical conditions or treatments that might lower seizure threshold should be evaluated in each patient.

Mianserin often produces drowsiness during the first weeks of treatment but has much less anticholinergic side-effects than other tricyclic antidepressants. It has less effects on producing hypotension and cardiac effects and there is only a low propensity for weight gain. Mirtazapine also has an increased amount of drowsiness and sedation. These side-effects are usually mild and transient. Mirtazapine has a low propensity to produce orthostatic hypotension or cardiac effects. There is a tendency for increased appetite and weight gain, however, which does not appear to be as severe as with tricyclics such as amitriptyline. Mianserin and mirtazapine have not been shown to produce high rates of sexual dysfunction as has been seen with trazodone and there is little evidence of lowering of the seizure threshold.

The side-effect profiles of the antidepressants are thought to relate to their respective effects on a variety of neurotransmitter systems. Clinicians should be aware of the profile of side-effects for each of the antidepressants they prescribe. The dose and duration of treatment interact with the intensity and type of side-effect and should be considered relative to antidepressant effects when evaluating, switching or stopping treatment. Although all antidepressant treatments can provoke switches into mania in vulnerable patients with bipolar disorder, it would appear that the MAOIs have a somewhat higher propensity to do this than the other compounds.

It is important that the nature of somatic and behavioural symptoms be carefully recorded before the onset of treatment so that the emergence of side-effects can be documented for the individual patient.

Toxic effects

There is ongoing concern recently regarding the issue of antidepressant use and suicide. The field has been grapplling with two inter-related issues: the possible risk of suicidal behaviour attributable to antidepressant treatment versus the potential decrease in suicidal behaviour afforded by antidepressant therapy. In 2004, there was an 18 per cent increase in adolescent suicides over the previous year.[43] This coincided with increased publicity about the relationship between antidepressant treatment and suicide risk in children and adolescents and a subsequent decline in antidepressant prescriptions. The Food and Drug Administration (FDA) has issued a black box warning to warn the public about the increased risk of suicidal thoughts and behaviour ('suicidality') in children and adolescents being treated with antidepressant medications.[44]

Often the most serious toxic effects are the result of overdose. Since depressed patients are at increased risk of suicide there is always the possibility that suicidally depressed patients will overdose on their antidepressants. This is a very serious consideration and should be carefully evaluated when prescribing antidepressants. The symptoms and course of events following acute antidepressant overdose are complex and can be confusing unless a clear history of overdose is obtained. With tricyclic antidepressants, restlessness and excitement are initially seen with possible myoclonus, and dystonia and seizures leading to the development of coma. Seriously compromized patients can have depressed respiration with hypoxia, depressed reflexes, hypertension, and hypothermia. With the antidepressants that have antimuscarinic activity, there can be strong anticholinergic effects with mydriasis, flushed skin, dry membranes, and tachycardia.

Antidepressant overdose can be life-threatening and patients should receive immediate emergency medical evaluation. The local poison control centre should be contacted in any case of suspected antidepressant overdose. Appropriate follow-up can include the use of activated charcoal to absorb the drug as well as other medical supportive measures. Different compounds have different probability of serious complications following an overdose and this is related to the amount ingested. However, drugs are often taken in combination, and it is difficult to know the exact composition and amount of the overdose.

Another serious toxic side-effect is the interaction of MAOIs and foods that are high in tyramine and other monoamines. Tyramine has both direct and indirect sympathomimetic actions, has a pressor action, and is present in a number of foodstuffs. It is normally broken down by the MAO enzymes and in the presence of a MAOI will increase in concentration. Some of the foods that should be restricted in the diet of patients taking MAOIs are listed in Table 6.2.3.7. The reaction usually develops 20 min to 1 h following ingestion of food and is characterized by nausea, apprehension, occasional chills, sweating, restlessness and hypotension with occipital headache, palpitations, and possibly vomiting. Neck stiffness, piloerection, dilated pupils, fever, and motor agitation are seen on examination. In severe forms, the reaction can lead to delirium, hyperpyrexia, cerebral hemorrhage, and death. The interaction of the irreversible MAOIs with certain dietary components leading to the hypertensive reaction is one of the most serious drawbacks to the use of these types of compounds. The reversible MAOIs moclobemide and brofaromine have not been found to interact with tyramine in the same fashion as the irreversible MAOIs. Thus, they have much less liability in terms of producing the hypertensive crisis seen with the irreversible MAOIs.[29] Phentolamine (5 mg) administered intravenously or nifedipine (onset of action 5 minutes), a calcium channel blocker, have been shown to be useful in the treatment of hypertensive reactions.

Serotonin syndrome is most often encountered when a MAOI is combined with an SSRI and there is an excess of 5-HT which overstimulates serotonin receptors.[45] This syndrome can manifest itself with sweating, diarrhoea, abdominal pain, fever, tachycardia, elevated blood pressure, myoclonus, hyper-reflexia, and with irritability and agitation. In its severe form, there can be severe hyperpyrexia, motor irritability, cardiovascular shock, and death. This toxic effect can result from the use of irreversible MAOIs and the addition of high amounts of tryptophan or other drugs that release serotonin in the brain. In addition, a common cause of this syndrome can result from the use of SSRIs and irreversible MAOIs concomitantly. Thus, it is strongly recommended that when MAOIs or SSRIs are utilized in sequence that the switch of treatment from one drug to the other has a minimum of a 14-day washout drug-free period before the second drug is started. In the case of discontinuing drugs with long half-lives such as fluoxetine an even longer period of up to 3 to 5 weeks may be necessary to safely avoid any possibility of producing the serotonin syndrome as a possible reaction to the drug combination.

The use of antidepressants during pregnancy is controversial and as always clinicians need to balance up the risks and benefits of treatments for individuals in this situation. Some recent studies have suggested that the use of SSRI medication in the perinatal period may be associated with adverse events like low birth weight and respiratory distress in the new born.[46, 47] However, this needs to be balanced against the fact that women at risk for depression may be at risk if not treated with antidepressants during pregnancy[48] and both these risks need to be balanced against each other.

Table 6.2.3.7 Dietary restrictions for patients on MAOIs

Aged cheeses	Liver	Raisins
American cheese[a]	Aged meats	Soy Sauce
Cottage cheese[a]	Canned meats	Ripe avocado
Yogurt[a]	Processed meats	Sauerkraut
Sour cream[a]	Meat extract	Licorice
Wine	Fermented foods	Chocolate[a]
Beer	Snails	Coffee[a]
Yeast extract	Anchovies	
Herring	Canned figs	
Sardines	Fava beans	

[a] Not over 50g daily.

The elderly are much more susceptible to toxic effects of antidepressants than younger individuals. In elderly patients, there may be other illnesses and the compensatory biological systems are not as resistant as in younger individuals. Mild toxic effects can be life threatening in the elderly. The newer antidepressants with fewer side-effects are the best drugs to use in the elderly.

There are a number of other toxic events that occur with less regularity. Isocarboxizid and phenelzine, since they are hydrazines, have some propensity to produce liver toxicity. Other much less frequent toxic events have occurred following antidepressants such as idiosyncratic individualized allergic reactions to the drug, suppression of the haematopoietic system, unusual dermatological reactions and hyponatremia.[49] There are reports of death in children receiving desipramine[50] though the cause of these deaths is unknown. Similarly, there is literature documenting increased incidence of gastric bleeding in association with SSRI treatment though there is a confounding effect of concomitant non-steroidal anti-inflammatory medication.[51]

Indications

Table 6.2.3.8 lists conditions where some antidepressant drugs have been found to be effective. Not all drugs are equally effective in each condition and very few clinical trials of the different compounds in each of the conditions have been conducted. Since the efficacy of antidepressant drugs is in part related to the dose administered and/or blood levels, it is difficult to be certain of the relative efficacy of one compound versus another when only single fixed doses are used. The expense and difficulty of multidose designs in comparing two treatments are extremely large and this is the main factor limiting comparisons of different drugs across the conditions listed in Table 6.2.3.8. In addition, the large number of compounds available would make this a very difficult task indeed. Another issue is that many of the drugs are only officially approved by the American Food and Drug Administration for use in depressed patients. Many of the indications listed in Table 6.2.3.8 are 'off-label' use of the medication. Since depression is the most prevalent illness, pharmaceutical companies have developed and brought forward drugs with depression as the primary indication. The expense of clinical trials to gain approval for other indications is high. Thus, for many of the conditions listed in Table 6.2.3.8, there is only fragmentary evidence for efficacy of some antidepressants and almost no data or comparable efficacy across drugs.

There are a number of different diagnostic approaches to depression as listed in Table 6.2.3.8. By and large, all of the drugs listed in Table 6.2.3.5 have been shown to be effective in the treatment of major depression. Most drugs have been studied in outpatient samples of patients with major depression. Their relative efficacy in the treatment of more severe conditions such as melancholia, psychotic depression, or bipolar depression remains limited. In addition, the relative efficacy of the different compounds as treatments for the depressive subtype, such as atypical depression, dysthymia, or secondary depression, has not been fully studied. There have been some reports that the MAOIs may be more effective in atypical depression[52] but not all studies have validated this. When depression in the elderly is under consideration, the side-effect profile for each drug becomes a much more relevant consideration when choosing a specific drug.

Anxiety disorders have considerable comorbidity with depression. Imipramine was initially found to be effective in the treatment of panic disorder and since then SSRIs have also been effective as well as MAOIs.[53] Clomipramine was found to be effective in obsessive-compulsive disorder though more recently SSRIs tend to be favoured as they generally have fewer side-effects. Depending on the studies, both SSRIs and MAOIs have been effective treatments in social phobia as well as some tricyclic drugs. In general anxiety disorder and post-traumatic stress disorder, antidepressants have also shown efficacy but not to the same extent as seen in panic disorder.[54]

It is of interest that some antidepressants have been effective in treating eating disorders. They are effective in bulimia nervosa[55] but not in anorexia nervosa.[56] The dose of fluoxetine to treat

Table 6.2.3.8 Clinical indications for antidepressant treatments

Clinical condition
Depression
Major depression
Melancholia
Psychotic depression
Bipolar depression
Atypical depression
Secondary depression
Dysthymia
Depression in elderly (pseudodementia)
Prevention of depression relapse
Anxiety disorders
Panic disorder
Obsessive–compulsive disorder
Social phobia
Generalized anxiety disorder
Post-traumatic stress disorder
Eating disorders
Bulimia nervosa
Obesity
Nausea with chemotherapy
Sleep disorders
Insomnia
Narcolepsy
Sleep apnea
Pain
Migraine headache
Atypical facial pain
Chronic pain syndromes
Diabetic neuropathy
Other disorders
Substance abuse
Alcoholism
Smoking cessation
Borderline personality disorder
Neurological disorders
Enuresis
Attention-deficit disorder
Premenstrual dysphoric disorder
Peptic ulcer
Urticaria pruritus
Premature ejaculation

bulimia nervosa is higher than the treatment of depression. The increase of weight seen following many antidepressants contrasts with some reports of the usefulness of SSRIs in the treatment of obesity.

In clinical practice, many clinicians have used trazodone as a night-time sedative. In the treatment of depression, sleep is one of the first symptoms to show improvement following initiation of most antidepressant treatments. Various reports of use of antidepressants in the treatment of narcolepsy and sleep apnoea have also been published.

Antidepressants have been effective in various pain syndromes. Since there is a wide range of the medical conditions producing pain, the results have been quite variable. In general the antidepressants have been able to reduce many of the painful symptoms as well as be effective in treating the secondary depression associated with chronic pain. However, they do not demonstrate the clear analgesic effect of drugs such as opioids.

Antidepressants have been reported to be effective in many other disorders including substance abuse, alcoholism, and smoking cessation.[57] In children with enuresis a dose of imipramine as low as 25 mg has been seen to be safe and effective. In both children and adults, imipramine, desipramine, bupropion, and nortriptyline have been effective in the treatment of attention-deficient disorder.[58] Antidepressants have found use in the treatment of premenstrual disorders,[59] and they are also useful in the treatment of several neurological disorders.

In general the indications and uses of a specific antidepressant in part depend on their side-effect profile and on the previously demonstrated efficacy. A major issue in the use of drugs to treat the large number of depressed patients with a comorbid medical condition is the careful choice of drug to minimize possible negative interactions with the medical disease.

Contraindications

The major contraindications in the use of antidepressants arise from the interaction of the pharmacological effects of antidepressant treatment with a comorbid condition of the patient or with diet or drug interactions. As mentioned above, the most serious contraindications arise from the use of irreversible MAOIs in patients taking other drugs or a diet that interacts and potentiates monoamine function resulting in a hypertensive crisis. A major contraindication is the use of MAOI in patients who receive anaesthesia.[59] Patients on MAOIs should carry a card for medical emergencies warning of drug interactions. Drugs that potentiate serotonin can interact with SSRIs to give the serotonin syndrome. The more relative contraindications involve the interaction of the side-effect profile of the antidepressant treatment with either the primary medical disease or with other medications that the patient may be taking. Another relative contraindication is the use of antidepressants during pregnancy and breast feeding though clearly there needs to be a risk benefit analysis of the use of the medication.

Drug interactions

Many patients may be taking other medications and many are prescribed more than one psychotrophic drug at a time. Because of this, it is important that clinicians are aware of drug-drug interactions. Drugs that impair the cytochrome P-450 microsomal enzyme system in the liver can interact with other drugs that are dependent on hepatic metabolism. For example, barbiturates and carbamazepine which induce hepatic enzymes can accelerate tricyclic metabolism and reduce steady state blood levels. Another anticonvulsant increasingly prescribed in the control of affective disorders, valproate, can reduce tricyclic drug clearance. Neuroleptics can elevate tricyclic blood levels which may be related to the impairment of the hydroxylation pathway for tricyclic metabolism. One of the more important drug-drug interactions involves the use of SSRIs and tricyclic drugs. This is related to the competitive inhibition of cytochrome by all of the SSRIs except fluvoxamine. This can result in clear elevations of steady state plasma concentrations. If combinations like this are utilized, the tricyclic doses need to be reduced. The utilization of a drug where plasma concentrations can be monitored (see Table 6.2.3.10) would help in the adjustment of dose if the tricyclic is combined with an SSRI.

One of the more serious drug-drug interactions previously mentioned is the interaction of tricyclic drugs with MAOIs. This can lead to hypertensive reactions and possibly stroke as well as possible induction of the serotonin syndrome.[45] Often antidepressants are combined with phenothiazines. There is some evidence that chlorpromazine can block the metabolism of tricyclics and thus when these two treatments are combined a possible reduction in the tricyclic treatment may be required. Other drugs that have been shown to increase tricyclic levels through blocking their metabolism include cimetidine, methylphenidate, and haloperidol. Tricyclic drugs can reduce the effects of clonidine and guanethidine in reducing blood pressure; an anticonvulsant, phenytoin, may be elevated; and the drug warfarin may be increased following tricyclic drugs. With the SSRIs, since there is a narrow pharmacological effect, interactions with anticholinergic agents or antihistaminics or alcohol are generally less than the tricyclics. The one major interaction is through the cytochrome P-450 family of enzymes which are inhibited by most SSRIs and interact with the metabolism of other drugs.

Some of the more serious drug interactions occur with the MAOIs. In addition to the dietary interactions, the MAOIs can interact with many of the 'over-the-counter' medications such as cough syrups and decongestants. Table 6.2.3.9 lists a number of compounds that have adverse drug interactions with MAOIs. Certainly many of the drugs that are direct or indirect adrenergic and dopaminergic agonists can produce overstimulation of the sympathetic nervous system. This can result in increased blood pressure and possibly adverse effects on the central nervous system. These drugs include all of the sympathomimetics, amphetamines, methylphenidate, and other stimulants. This can also occur with drugs such as other MAOIs and tricyclics or SSRIs that increase monoamine levels. MAOIs may worsen hypoglycaemia and require readjustment of the dosage of hypoglycaemic agents. Major concerns arise when patients on MAOIs need surgery because of the interaction with a number of compounds used in anaesthesia. This is more likely to occur with the use of pethidine. Careful consideration should be given to using a minimal 2-week washout for patients on MAOIs under going elective surgery.

The mixed 5-HT agonist drugs also have important drug-drug interactions. Trazodone can potentiate barbiturate and alcohol and it can increase drowsiness and sedation in patients taking these

Table 6.2.3.9 Adverse drug interactions with MAOIs

Other antidepressants	Other MAOIs
Buspirone	Carbamazepine
Stimulants	L-Dopa
Sympathomimetics	Methyldopa
Dopamine	Guanethidine
Amphetamines	Dextromethorphan
Methylphenidate	Pethidine
Adrenaline	Cocaine
Asthma inhalants	Reserpine
Decongestants	Tryptophan
Appetite suppressants	Fenfluramine

agents. It has also been reported to produce the serotonin syndrome at times when combined with an SSRI and possibly buspirone. Trazodone has altered the kinetics of benzodiazepines including alprazolam and triazolam. Trazodone interacts with cytochrome P-450 enzymes and has been shown to increase plasma levels of digoxin. Unlike trazodone, nefazodone does not appear to potentiate the sedative effects of alcohol.

Bupropion undergoes hepatic metabolism and its levels can be altered by other drugs effecting this metabolic route. There is some dopaminergic activation with bupropion and it has had adverse interactions with MAOIs. Because of the dopaminergic activity, there have been interactions with anti-parkinsonian medication. Because bupropion lowers the seizure threshold at higher dosages it can interact and with other medications that would have similar effects to produce seizures. A combination of bupropion and lithium may increase the likelihood of seizures. There are some reports that carbamazepine may decrease bupropion drug levels.

Mirtazapine is metabolized by cytochrome P-450 enzyme systems. There is the potential for interaction with other drugs via this system. The extent of use of mirtazapine is not as great as the older drugs and the drug-drug interactions are not as extensively reported.

In general, a large number of drug-drug interactions have been reported for the antidepressants. The drug-drug interactions can be quite variable depending on the patient and the dosage and duration of treatments. Thus, the adverse drug-drug interactions are one of the main reasons for the recommendations to use monotherapy rather than more than one drug. The use of drug combinations should, on average, be restricted to patients who have a poor response to a single treatment because of the possibility of adverse drug-drug interactions.

Drug withdrawal

Long-term administration of a drug can produce adaptive changes in many aspects of the human biology. When the drug is abruptly discontinued, 'rebound' drug withdrawal symptoms can be observed. This is most clearly seen with longer-term opiate, benzodiazepine, or barbiturate use. Antidepressants are not addictive and dependence does not develop. With the antidepressants some degree of tolerance to the sedative and autonomic effects tends to develop and on abrupt withdrawal patients can have emerging symptoms consisting of malaise, dizziness, nausea, diarrhoea, chills, insomnia, restlessness, and muscle aches. Symptoms emerging during drug withdrawal have been seen following treatment with tricyclics as well as SSRIs.[60] They have been described on occasion for MAOIs and the other non-monoamine uptake antidepressants. One main factor is the drug elimination half-life. Abrupt discontinuation of drugs with a short elimination half-life will produce more emergent side-effects than drugs with a long half-life. Thus, it has been found that patients taking shorter half-life drugs such as paroxetine and sertraline have more of an emergent symptom increase than patients taking the longer elimination half-life drug fluoxetine. Therefore, the dose of drugs with a shorter elimination half-life should generally be tapered over a 2- to 3-week period when being discontinued rather than being stopped abruptly. Drugs with a longer elimination half-life can be stopped abruptly since the parent drug and metabolite may last for many days. Clinicians must be aware that slow elimination means that parent drug and active metabolites remain in the body for up to several weeks.

With the irreversible MAOI inhibitors, since there is a 1- or 2-week period during with monoamine oxidase must recover following discontinuation of the MAOI, emergent side-effects have not been as regularly observed. In general with the other antidepressants the withdrawal syndromes have not been permanent. Clinicians must make special efforts to discriminate between the return of symptoms and the emergence of new symptoms related to drug withdrawal.

Dosages and administration

In Table 6.2.3.10 some of the suggested optimal plasma concentrations for different antidepressant drugs are listed. For nortriptyline, desipramine, imipramine, and amitriptyline there is some evidence for a minimal plasma and concentration necessary for clinical response. An established therapeutic range is available for nortriptyline. Thus, patients with nortriptyline concentrations between 50 and 150 mg/ml seem to do much better. With the other drugs, it is generally thought that the plasma levels reflect a minimal threshold of plasma concentration for clinical response. Below this level patients are less likely to respond and the upper limit indicates that there is increased possibility of the systemic or cardiac toxicity. Bupropion levels between 50 and 100 ng/ml may possibly be the best range. However, it is important that the dose be kept below 450 mg/day because of the possibility of seizures.

Except for nortriptyline and the use of plasma concentrations to obtain a minimal effective level, it is generally the patient's clinical response that dictates dosage adjustments. One difficulty is that some patients with plasma concentrations outside the therapeutic range do respond and many patients with concentrations within the therapeutic range do not. Thus, the dosage needs to be adjusted depending on the individual patient's response. Clearly plasma concentration monitoring can be helpful in many situations such as evaluating plasma levels when higher than standard doses are used, assessing toxicity, use in elderly patients or patients with comorbid conditions to evaluate possible drug interactions, or where compliance is questioned. Blood for drug levels is usually obtained for plasma levels during elimination phase which is usually in the morning 12 h after the last dose.

Table 6.2.3.10 Dosage and administration of antidepressant drugs. Doses in brackets refer to doses recommended for elderly patients

Drug	Initial dose (mg/day)	Theraupeutic dose range (mg/day)	Recommended optimal plasma concentration (ng/ml)
Monoamine reuptake inhibitors			
Imipramine	25–75 (10)	150–300 (30–50)	200–300[d]
Desipramine	50–75	100–300 (25–150)	125–300[d]
Amitriptyline	75 (30–75)	150–200	120–250[d]
Nortriptyline	25	75–150	50–150
Trimipramine	50–75 (30–75)	150–300 (75–150)	NA
Clomipramine	10	30–250 (30–75)	NA
Protriptyline	20–40 (15)	20–60	70–240
Doxepin	75 (10–50)	30–300 (30–50)	110–250[d]
Amoxapine	100–150 (50–75)	100–600 (150–300)	200–400[e]
Maprotiline	25–75 (25)	25–225 (50–75)	200–300[d]
Venlafaxine	75	150–375	NA
Milnacipran	50	50–100	NA
Duloxetine	60	60	NA
Reboxetine	8	10–12	NA
SSRIs			
Fluoxetine	20	20–60	NA
Sertraline	50	50–200	NA
Fluvoxamine	50–100	100–300	NA
Paroxetine	20	50 (40)	NA
Citalopram	20	20–60 (20–40)	NA
Escitalopram	10 (5)	10–20 (10–20)	NA
MAOIs			
Isocarboxazid	30	10–60 (5–10)	NA
Phenelzine	45	15–90	NA
Tranylcypromine	20	10–30	NA
Moclobemide	300	150–600	NA
Brofaromine	50	50–150	NA
Selegiline	6[a]	6–12[a]	NA
Other antidepressants			
Trazodone	100 (150)	150–600	NA
Nefazodone	200 (100)[b]	200–600[b]	NA
Buproprion	200[c]	300–450[c]	50–100
Mianserin	30–40 (30)	30–90	NA
Mirtazapine	15	45	NA

NA, data not available.

[a] Selegiline not licensed in UK for depression. Doses quoted refer to transdermal patches marketed in the US.

[b] Nefazodone not licensed in UK.

[c] Buproprion not licensed in Uk for depression.

[d] Parent drug plus demethylated metabolite.

[e] Parent drug plus hydroxymetabolite.

These doses should not be used for patient prescriptions; clinicians should consult manufacturer's literature for recommendation of doses and frequency of administration.

Acknowledgement

The authors would like to thank Aybala Saricicek MD for her assistance in the preparation of the chapter.

Further information

Biological Psychiatry (Elsevier)—the journal of the Society of Biological Psychiatry http://www.sobp.org/

Neuropsychopharmacology (Nature Publishing group)—the journal of the American College of Neuropsychopharmacology http://www.acnp.org/

Schatzberg, A.F. and Nemeroff, C. B. (eds). (2004). *American Psychiatric Publishing Textbook of Psychopharmacology* (3rd edn.) American Psychiatric Publishing, Arlington, VA.

References

1. Lopez, A.D., Murray, C.C.J.L. (*1*998). The global burden of disease, 1990-2020. *Nature Medicine,* **4**, 1241–3.
2. Bloch, R.G., Doonief, A.S., Buchberg, A.S., *et al.* (1954). The clinical effect of isoniazid and iproniazid in the treatment of pulmonary tuberculosis. *Annals of Internal Medicine,* **40**, 881–900.
3. Crane, G.E. (1957). Iproniazid (marsilid) phosphate, a therapeutic agent for mental disorders and debilitating diseases. *Psychiatric research reports,* **135**, 142–52.
4. Kuhn, R. (1958). The treatment of depressive states with G 22355 (imipramine hydrochloride). *American Journal of Psychiatry,* **15**, 459–64.
5. Coppen, A., Shaw, D.M., Farrell, J.P. (1963). Potentiation of the antidepressive effect of a monoamine-oxidase inhibitor by tryptophan. *Lancet*, **1**, 79–81.

6. Schildkraut, J.J. (1965). The catecholamine hypothesis of affective disorders: a review of supporting evidence. *American Journal of Psychiatry,* **122**, 509–22.
7. Bunney, W.E., Jr., Davis, J.M. (1965). Norepinephrine in depressive reactions. A review. *Archives of General Psychiatry,* **13**, 483–94.
8. Vetulani, J., Sulser, F. (1975). Action of various antidepressant treatments reduces reactivity of noradrenergic cyclic AMP-generating system in limbic forebrain. *Nature,* **257**, 495–6.
9. Charney, D.S., Menkes, D.B., Heninger, G.R. (1981). Receptor sensitivity and the mechanism of action of antidepressant treatment. Implications for the etiology and therapy of depression. *Archives of General Psychiatry,* **38**, 1160–80.
10. Blier, P., de Montigny, C., Chaput, Y. (1987). Modifications of the serotonin system by antidepressant treatments: implications for the therapeutic response in major depression. *Journal of Clinical Psychopharmacology,* **7**, 24S–35S.
11. Paul, I.A., Nowak, G., Layer, R.T., *et al.* (1994). Adaptation of the N-methyl-D-aspartate receptor complex following chronic antidepressant treatments. *Journal of Pharmacology and Experimental Therapeutics,* **269**, 95–102.
12. Layer, R.T., Popik, P., Olds, T., *et al.* (1995). Antidepressant-like actions of the polyamine site NMDA antagonist, eliprodil (SL-82.0715). *Pharmacology, Biochemistry, and Behaviour*, **52**, 621–7.
13. Holsboer, F., Barden, N. (1996). Antidepressants and hypothalamic-pituitaryadrenocortical regulation. *Endocrine Reviews*, **17**, 187–205.
14. Griebel, G., Perrault, G., Sanger, D.J. (1998). Characterization of the behavioural profile of the non-peptide CRF receptor antagonist CP-154,526 in anxiety models in rodents. Comparison with diazepam and buspirone. *Psychopharmacology (Berl),* **138**, 55–66.
15. Duman, R.S., Heninger, G.R., Nestler, E.J. (1997). A molecular and cellular theory of depression. *Archives of General Psychiatry,* **54**, 597–606.
16. Wilde, M.I., Benfi eld, P. Tianeptine. (1995). A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic effi cacy in depression and coexisting anxiety and depression. *Drugs,* **49**, 411–39.
17. Kramer, M.S., Cutler, N., Feighner, J., *et al.* (1998). Distinct mechanism for antidepressant activity by blockade of central substance P receptors. *Science,* **281**, 1640–5.
18. Taylor, M.J., Freemantle, N., Geddes, J.R., *et al.* (2006). Early Onset of Selective Serotonin Reuptake Inhibitor Antidepressant Action: Systematic Review and Meta-analysis. *Archives of General Psychiatry,* **63**, 1217–23.
19. Berman, R.M., Cappiello, A., Anand, A. *et al.* (2000). Antidepressant effects of ketamine in depressed patients. *Biological Psychiatry,* **47**, 351–4.
20. Zarate, C.A., Jr., Singh, J.B., Carlson, P.J. *et al.* (2006). A Randomized Trial of an N-methyl-D-aspartate Antagonist in Treatment-Resistant Major Depression. *Archives of General Psychiatry,* **63**, 856–64.
21. Maeng, S., Zarate, C.A., Jr., Du, J. *et al.* (207). Cellular Mechanisms Underlying the Antidepressant Effects of Ketamine: Role of alpha-Amino-3-Hydroxy-5-Methylisoxazole-4-Propionic Acid Receptors. *Biological Psychiatry,* **63**, 349–52.
22. Richelson, E. (2003). Interactions of antidepressants with neurotransmitter transporters and receptors and their clinical relevance. *Journal of Clinical Psychiatry,* **64** Suppl **13**, 5–12.
23. Gutierrez, M.A., Stimmel, G.L., Aiso, J.Y. (2003). Venlafaxine: a 2003 update. *Clinical Therapeutics,* **25**, 2138–54.
24. Stahl, S.M., Grady, M.M., Moret, C., *et al.* (2005). SNRIs: their pharmacology, clinical efficacy, and tolerability in comparison with other classes of antidepressants. *CNS Spectrums,* **10**, 732–47.
25. Preskorn, S.H. (2004). Reboxetine: a norepinephrine selective reuptake pump inhibitor. *Journal of Psychiatric Practice,* **10**, 57–63.
26. Hyttel, J., Bogeso, K.P., Perregaard, J., *et al.* (1992). The pharmacological effect of citalopram residues in the (S)-(+)-enantiomer. *Journal of Neural Transmission. General Section,* **88**, 157–60.
27. Kennedy, S.H. (1997). Continuation and maintenance treatments in major depression: the neglected role of monoamine oxidase inhibitors. *Journal of Psychiatry & Neuroscience,* **22**, 127–31.
28. Patkar, A.A., Pae, C.U., Masand, P.S. (2006). Transdermal selegiline: the new generation of monoamine oxidase inhibitors. *CNS Spectrums,* **11**, 363–75.
29. Robinson, D.S. (2002). Monoamine oxidase inhibitors: a new generation. *Psychopharmacology Bulletin,* **36**, 124–38.
30. Haria, M., Fitton, A., McTavish, D. (1994). Trazodone: A review of its pharmacology, therapeutic use in depression and therapeutic potential in other disorders. *Drugs & Aging,* **4**, 331–55.
31. Dunner, D.L., Laird, L.K., Zajecka, J., *et al.* (2002). Six-year perspectives on the safety and tolerability of nefazodone. *Journal of Clinical Psychiatry,* 63 Suppl **1**, 32–41.
32. Ascher, J.A., Cole, J.O., Colin, J.N. *et al.* (1995). Bupropion: a review of its mechanism of antidepressant activity. *Journal of Clinical Psychiatry,* **56**, 395–401.
33. Leverich, G.S., Altshuler, L.L., Frye, M.A. *et al.* (2006). Risk of Switch in Mood Polarity to Hypomania or Mania in Patients With Bipolar Depression During Acute and Continuation Trials of Venlafaxine, Sertraline, and Bupropion as Adjuncts to Mood Stabilizers. *American Journal of Psychiatry,* **163**, 232–9.
34. Hurt, R.D., Sachs, D.P., Glover, E.D. *et al.* (1997). A comparison of sustained-release bupropion and placebo for smoking cessation. *New England Journal of Medicine,* **337**, 1195–202.
35. Szegedi, A., Schwertfeger, N. (2005). Mirtazapine: a review of its clinical efficacy and tolerability. *Expert Opinion on Pharmacotheraphy,* **6**, 631–41.
36. Hemeryck, A., Belpaire, F.M. (2002). Selective serotonin reuptake inhibitors and cytochrome P-450 mediated drug-drug interactions: an update. *Current Drug Metabolism,* **3**,13–37.
37. Kapur, S., Cho, R., Jones, C., *et al.* (1999). Is amoxapine an atypical antipsychotic? Positron-emission tomography investigation of its dopamine2 and serotonin2 occupancy. *Biological Psychiatry,* **45**, 1217–20.
38. Scates, A.C., Doraiswamy, P.M. (2000). Reboxetine: a selective norepinephrine reuptake inhibitor for the treatment of depression. *Annals of Pharmacotheraphy,* **34**, 1302–12.
39. Detke, M.J., Lu, Y., Goldstein, D.J., *et al.* (2002). Duloxetine, 60 mg once daily, for major depressive disorder: a randomized double-blind placebo-controlled trial. *Journal of Clinical Psychiatry,* **63**, 308–15.
40. Sussman, N., Ginsberg, D.L., Bikoff, J. (2001). Effects of nefazodone on body weight: a pooled analysis of selective serotonin reuptake inhibitor- and imipramine-controlled trials. *Journal of Clinical Psychiatry,* **62**, 256–60.
41. Spencer, T., Biederman, J., Steingard, R., *et al.* (1993). Bupropion exacerbates tics in children with attention-deficit hyperactivity disorder and Tourette's syndrome. *Journal of the American Academy of Child Adolescent Psychiatry,* **32**, 211–4.
42. Alper, K., Schwartz, K.A., Kolts, R.L., *et al.* (2007). Seizure Incidence in Psychopharmacological Clinical Trials: An Analysis of Food and Drug Administration (FDA) Summary Basis of Approval Reports. *Biological Psychiatry*, **62**, 345–54.
43. Hamilton, B.E., Minino, A.M., Martin, *et al.* (2007). Annual summary of vital statistics: 2005. *Pediatrics,* **119**, 345–60.
44. Hammad, T.A., Laughren, T., Racoosin, J. (2006). Suicidality in Pediatric Patients Treated With Antidepressant Drugs. *Archives of General Psychiatry,* **63**, 332–9.
45. Boyer, E.W., Shannon, M. (2005). The serotonin syndrome. *New England Journal of Medicine, 352*, 1112–20.
46. Chambers, C.D., Hernandez-Diaz, S., Van Marter, L.J. *et al.* (2006). Selective serotonin-reuptake inhibitors and risk of persistent pulmonary hypertension of the newborn. *New England Journal of Medicine,* **354**, 579–87.

47. Oberlander, T.F., Warburton, W., Misri, S., *et al.* (2006). Neonatal Outcomes After Prenatal Exposure to Selective Serotonin Reuptake Inhibitor Antidepressants and Maternal Depression Using Population-Based Linked Health Data. *Archives of General Psychiatry,* **63**, 898–906.
48. Cohen, L.S., Altshuler, L.L., Harlow, B.L. *et al.* (2006). Relapse of Major Depression During Pregnancy in Women Who Maintain or Discontinue Antidepressant Treatment. *Journal of the American Medical Association,* **295**: 499–507.
49. Madhusoodanan, S., Bogunovic, O.J., Moise, D., *et al.* (2002). Hyponatraemia associated with psychotropic medications. A review of the literature and spontaneous reports. *Adverse Drug Reactions and Toxicological Reviews,* **21**, 17–29.
50. Amitai, Y., Frischer, H. (2006). Excess fatality from desipramine in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry,* **45**, 54–60.
51. Yuan, Y., Tsoi, K., Hunt, R.H. (2006). Selective serotonin reuptake inhibitors and risk of upper GI bleeding: confusion or confounding? *American Journal of Medicine,* **119**, 719–27.
52. Quitkin, F.M., Stewart, J.W., McGrath, P.J., *et al.* (1993). Columbia atypical depression. A subgroup of depressives with better response to MAOI than to tricyclic antidepressants or placebo. *British Journal Psychiatry Supplement,* 30–4.
53. Otto, M.W., Tuby, K.S., Gould, R.A., *et al.* (2001). An effect-size analysis of the relative efficacy and tolerability of serotonin selective reuptake inhibitors for panic disorder. *American Journal of Psychiatry,* **158**, 1989–92.
54. Ball, S.G., Kuhn, A., Wall, D., *et al.* (2005). Selective serotonin reuptake inhibitor treatment for generalized anxiety disorder: a double-blind, prospective comparison between paroxetine and sertraline. *Journal of Clinical Psychiatry,* **66**, 94–9.
55. Group FBNCS. (1992). Fluoxetine in the treatment of bulimia nervosa. A multicenter, placebo-controlled, double-blind trial. Fluoxetine Bulimia Nervosa Collaborative Study Group. *Archives of General Psychiatry,* **49**, 139–47.
56. Attia, E., Haiman, C., Walsh, B.T., *et al.* (1998). Does fluoxetine augment the inpatient treatment of anorexia nervosa? *American Journal of Psychiatry,* **155**, 548–51.
57. Nunes, E.V., Levin, F.R. (2004). Treatment of depression in patients with alcohol or other drug dependence: a meta-analysis. *Journal of the American Medical Association,* **291**, 1887–96.
58. Culpepper, L. (2006). Primary care treatment of attention-deficit/hyperactivity disorder. *Journal of Clinical Psychiatry,* **67 Suppl 8**, 51–8.
59. Yonkers, K.A. (2004). Management strategies for PMS/PMDD. *Journal of Family Practice,* Suppl: S15–20.
60. Ditto, K.E. (2003). SSRI discontinuation syndrome. Awareness as an approach to prevention. *Postgraduate Medicine*, **114**, 79–84.

6.2.4 Lithium and related mood stabilizers

Robert M. Post

Introduction

Global categorization of the mood stabilizers

Lithium is the paradigmatic mood stabilizer. It is effective in the acute and prophylactic treatment of both mania and, to a lesser magnitude, depression. These characteristics are generally paralleled by the widely accepted anticonvulsant mood stabilizers valproate, carbamazepine (Table 6.2.4.1), and potentially by the less well studied putative mood stabilizers oxcarbazepine, zonisamide, and the dihydropyridine L-type calcium channel blocker nimodipine. In contrast, lamotrigine has a profile of better antidepressant effects acutely and prophylactically than antimanic effects.

Differential responsivity among individual patients

Having grouped lithium, valproate, and carbamazepine together, it is important to note they have subtle differences in their therapeutic profiles and differential clinical predictors of response (Table 6.2.4.1). **Response to one of these agents is not predictive of either a positive or negative response to the others.**[1,2] Thus, clinicians are left with only rough estimates and guesses about which drug may be preferentially effective in which patients. Only sequential clinical trials of agents either alone or in combination can verify responsivity in an individual patient.[3] **Individual response trumps FDA-approval**.

Requirements for method of longitudinal assessment

Given this clinical conundrum, it is advisable that patients, family members, clinicians, or others carefully rate patients on a longitudinal scale in order to most carefully assess responses and side effects. These are available from the Depression Bipolar Support Alliance (DBSA), the STEP-BD NIMH Network, or www.bipolarnetworknews.org and are highly recommended.

Increasing need for complex combination treatment

The importance of careful longitudinal documentation of symptoms and side effects is highlighted by the increasing use of multiple drugs in combination.[2] This is often required because patients may delay treatment-seeking until after many episodes, and very different patterns and frequencies of depressions, manias, mixed states, as well as multiple comorbidities may be present. Treating patients to the new accepted **goal of remission** of their mood and other anxillary symptoms usually requires use of several medications. If each component of the regimen is kept below an individual's side-effects threshold, judicious use of multiple agents can reduce rather than increase the overall side-effect burden.

Patients and family education is a must

There is increasing evidence of reliable abnormalities of biochemistry, function, and anatomy in the brains of patients with bipolar disorder, and some of these are directly related to either duration of illness or number of episodes.[4,5] Therefore, as treatment resistance to most therapeutic agents is related to number of prior episodes, and brain abnormalities may also increase as well, it behooves the patient to begin and sustain acute and **long-term treatment as early as possible**.

Early age of onset and treatment delay are related to an adverse outcome in adulthood

Despite the above academic, personal, and public health recommendations, bipolar disorder often takes ten years or more to diagnose and, hence, treat properly. In fact, a **younger age** of onset is highly related to presence of a **longer delay** from illness onset to **first treatment, and** as well, to **a poorer outcome** assessed both retrospectively and prospectively.[6]

Table 6.2.4.1 Neuroprotective effects of lithium in cultured cells and animal models of diseases

I. Therapeutic Spectrum	Lithium	Carbamazepine	Valproate	Lamotrigine	Nimodipine
A Acute Mania	++++	++++	++++	0	(++)
B. Mania Prophylaxis	++++	+++	++++	+	(++)
C. Acute Depression	++	++	++	+++	(+)
D. Depression Prophylaxis	++++	++++	+++	++++	(++)
II. Correlates of Response					
A. Family History + = positive – = negative	+ Mania –Depression	– Bipolar Disorder	ND	+ Anxiety Disorder	ND
B. Bipolar Type (BP)	I	II	I,II	I,II	I,II,NOS
Manic Type: Pattern:	Euphoric Intermittent	Dysphoric Continuous	Dysphoric Non-accelerating	Continuous	Euphoric Ultradian
C. Comorbidities					
Anxiety Disorder	None	++	+++	+++	(+)
Substance Use	None	++	+++	+/–	(+)
D. PTSD Utility	0	++	++	++	0
E. Other Unique Targets	M-D-I vs. D-M-I Antisuicidal AntimedicalMortality	Alcohol Withdrawal Tigeminal Neuralgia	Alcohol Abstinence Migraine Prophylaxis		(Alzheimer's Dementia) (Migraine Prophylaxis) Subarachnoid Hemorrhage
F. Baseline PET Activity	?	Hyper- metabolism	?	Hypo-metabolism	Hypo-metabolism
G. CSF SRIF (Somatostatin)					
Predicts Response:	?	No Prediction	?	?	Low SRIF
Effect of Drug:	?	Decreases SRIF			Increases SRIF
III. Neurotropic Effects					
A. Increase BDNF	Yes	Yes	Yes	?	?
B. Increase Neurogenesis	Yes	?	Yes	?	?
C. Neuroprotective	Yes	Yes	Yes	Yes	(Yes)

Legend: ++++ = very marked; +++ = marked; ++ = moderate; + = mild or some effect; +/– = equivocal; () = ambiguous data; 0 = no effect; – = worse.

BDNF: A role in vulnerability, onset, progression, and treatment

New data indicate that the brain growth factor BDNF (brain-derived neurotrophic factor) which is initially important to synaptogenesis and neural development, and later neuroplasticity and long-term memory in the adult is involved in all phases of bipolar disorder and its treatment.[7]

It appears to be: 1) both a genetic (the val-66-val allele of BDNF) and environmental (low BDNF from childhood adversity) *risk factor*; 2) *episode-related* (serum BDNF decreasing with each episode of depression or mania in proportion to symptom severity; 3) related to some *substance abuse* comorbidity (BDNF increases in the VTA with defeat stress and cocaine self-administration); and 4) related to *treatment*. **Lithium, valproate, and carbamazepine increase BDNF and quetiapine and ziprasidone block the decreases in hippocampal BDNF that occur with stress (as do antidepressants).**

More episodes convey more problems

A greater number of prior episodes is related to increased likelihood of: 1) a rapid cycling course; 2) more severe depressive symptoms; 3) more disability; 4) more cognitive dysfunction; and 5) even the incidence of late life dementia.[4,8,9,10]

Early effective treatment may protect the brain

Taken together, the new data suggest a new view not only of bipolar disorder, but its treatment. Adequate effective **treatment may** not only (a) prevent affective episodes (with their accompanying risk of morbidity, dysfunction, and even death by suicide or the increased medical mortality associated with depression), but may also (b) **reverse or prevent some of the biological abnormalities associated with the illness from progressing.**

Thus, patients should be given timely information pertinent to their stage of illness and recovery that emphasizes not only the risk of treatments, but also their potential, figuratively and literally, life-saving benefits. Long-term treatment and education and targeted psychotherapies are critical to a good outcome.

Therapeutic strategy

We next highlight several attributes of each mood stabilizer, but recognize that the choice of each agent itself is based on inadequate information from the literature, and sequencing of treatments and their combinations is currently more an art than an evidence-based science. We look forward to these informational and clinical trial deficits being reduced in the near future and the development of single nucleotide polymorphism (SNP) and other neurobiological predictors of individual clinical response to individual drugs.

In the meantime, patients and clinicians must struggle with treatment choice based on: 1) the most appropriate targetting of the predominant symptom picture with the most likely effective agent (Table 6.2.4.1 and 6.2.4.2) the best side-effects profile for that patient (Table 6.2.4.2 and 6.2.4.3) using combinations of drugs with different therapeutic targets and mechanisms of action

Table 6.2.4.2 Global putative mechanisms of action

	Li	CBZ	VPA	LTG	NIMOD
Antiglutamineric:	+	+	+	+	?
Via:					
Glutamate Uptake	+				
Sodium Blockade		+	(+)	+	
↑ Brain GABA	+	+/–	++	—	0
↑ GABA-B R in hippocampus (with chronic administration)	+	+	+		
↓ Calcium Influx	+	+	+	+	++
Via:					
Weak NMDA Receptor Inhibition	+	+	+	+	0
Inhibition of Calcium-Channel Type	—	(L)	T	(N,P)	L
↓ DA Turnover	+	+	+		
Second Messenger System					
↓ c-AMP, G proteins	++	++	—	—	—
PI Turnover	↓	(↑)	N.C.	?	+/–
PKC Inhibition	++		++		
↓ ras, MEK, Erk Pathway	++		++		
↓ Inositol Transport	+	+	+	?	?
↑ BDNF	+	+	+		
↑ Bcl-2	+		+		
Histone Deacetylase Inhibition	—	+	++	?	?

Table 6.2.4.3 Global assessment of relative side-effects

	Li	CBZ	VPA	LTG	NIMOD
Weight Gain	++	+	++	0	0
Tremor	++	+/–	++	+/–	0
GI Upset	++	+	++	+/–	0
Memory Disturbance	+	+	+	+	–
Rash	0[a]	++	+/–	++	0
↓WBC	—	++	0	0	0
Agranulocytosis	0	+	0	0	0
↓Platelets	–	0[b]	++	0	0
↑ Liver Enzymes	0	++	++	(+)	0
Hepatitis	0	+	+	?	0
Dizziness Ataxia Diplopia	+/–	++	+	+	+/–
Hyponatremia	–	++	0	0	0
Alopecia	+/–	0	++	?	0
Thyroid Decrements	++	+/–	+/–	+/–	0
Teratogenic	+	++	++	0	0
Malformations	Epstein's Anomaly	Spina Bifida	Spina Bifida		
Developmental Delay	0	+/–	++	+/–	0

Legend: a = psoriasis; b=with aplastic anemia; ++ = moderate to substantial or common; + = mild to less frequent; +/– = equivocal or rare; 0 = not present or no change; – = opposite effect; () = ambiguous findings

(Table 6.2.4.3 and 6.2.4.4) careful consideration of potential advantageous pharmacodynamic interactions and disadvantageous pharmacokinetic drug-drug interactions that need to be avoided or anticipated.

Mood stabilizers

Lithium

In the late 1960s and early 1970s, open and double-blind randomized treatment and discontinuation studies revealed **highly significant effects of lithium in long-term prophylaxis**. These studies followed shortly after a series of studies demonstrating lithium's acute antimanic effects in comparison with both placebo and the existing neuroleptic treatments. High rates of response were touted and lithium clinics were established with the hope that the 60 to 80 per cent response rates revealed in the controlled clinical trials would be mirrored by clinical practice.[11]

However, over the past two decades there has been increasing recognition of the inadequacy of lithium both in acute treatment and long-term prophylaxis, even when used with adjunctive treatments such as antipsychotics, benzodiazepines, and antidepressants.[12,13,14] Given this increasing cognizance of lithium's less than dramatic efficacy in many patients with bipolar illness, alternative and adjunctive treatments were sought.

The anticonvulsants carbamazepine, valproate, and lamotrigine

Carbamazepine and valproate are now well recognized as potential mood-stabilizing anticonvulsants, and initial promising data are emerging for lamotrigine as well. The data are more equivocal for oxcarbazepine and minimal for zonisamide. Importantly, **the GABA-active anticonvulsants gabapentin, tiagabine, and topiramate** are **not effective in acute mania** and thus cannot be considered mood stabilizers. Nonetheless, topiramate may be useful in some common comorbidities of bipolar illness including alcohol and cocaine abuse, bulimia, overweight, migraine, and PTSD. Similarly, gabapentin (which increases brain GABA and acts at the alpha$_2$ delta subunit of the L-type calcium channel) and its close relative pregabalin may have secondary utility in comorbid panic anxiety and social phobia, sleep disorder, alcohol withdrawal, and chronic pain syndromes.

L-type calcium-channel blockers

The dihydropyridine L-type calcium channel blockers became a focus of possible interest for lithium-intolerant and lithium-unresponsive patients based on a variety of clinical and theoretical rationales. Dubovsky *et al.*[15] found increased intracellular calcium in blood elements of patients with bipolar illness, a finding that has been replicated more than a dozen times. Dubovsky *et al.*[16] proceeded to demonstrate the potential antimanic efficacy of the L-type calcium channel blocker verapamil. Many other small

controlled studies were also positive, although less than dramatic results have recently been reported by two groups.[17,18] In addition, one controlled study found that verapamil was not an effective acute antidepressant, even though it appeared to have good antimanic properties.[19] Verapamil was never widely used in clinical practice.

Given these ambiguities with verapamil, Pazzaglia *et al.*[20] and Post *et al.*[21] at the NIMH in the U.S. chose to explore the potential antimanic and antidepressant effects of the dihydropyridine L-type calcium channel blocker nimodipine which has a very different biophysical and pharmacological profile from verapamil.[22] In contrast to the phenylalkylamine verapamil, the dihydropyridine nimodipine is a more potent anticonvulsant that blocks cocaine-induced hyperactivity, sensitization, and dopamine overflow; is positive in animal models of depression; and it improves rather than impairs cognition.

Pharmacology

Lithium

A series of sequential decade-related candidate mechanisms for lithium's psychotropic (mood-stabilizing) actions have been suggested over the last 50 years. These included:

- Effects on enzymes and biosynthetic aminergic neurotransmitter pathways (1950s)
- Effects on presynaptic adrenergic release and reuptake mechanisms (1960s)
- Effects on postsynaptic receptor modulation and impact on receptor supersensitivity (1970s)
- Effects on second messenger systems, adenylate cyclase and G-proteins (1980s)[23]
- Effects on phosphoinositol turnover, protein kinases (PKC and GSK-36), and other signal transduction pathways (1990s)[24,25]
- Most recently, effects on the nuclear level of DNA binding and gene transcription have been found that increase neurotrophic and other factors regulating neuroprotection versus apoptosis and cell death (2000s)[26,27]

Several of these candidate mechanisms are summarized in Figs 6.2.4.1 and 6.2.4.2, and are compared with the mechanisms of some of the other mood-stabilizing anticonvulsants such as carbamazepine and valproate. Lithium's effects on G proteins[20] are conceptually intriguing in relation to lithium's ability to modulate overactive systems, and preliminary support for its adenylate cyclase effects being important for acute mania are based on the finding that novel adenylate cyclase inhibitors are also effective in acute mania.[21] Similarly, recent studies implicating the ability of lithium and valproate to inhibit protein kinase C has been preliminarily validated with the demonstration of antimanic effects of the protein kinase C inhibitor tamoxifen in two double-blind studies.[28,29]

Most recently, lithium has emerged as a possible **neurotrophic and neuroprotective drug** via multiple potential pathways. It inhibits calcium influx through the N-methyl-$_D$-aspartate glutamate receptor;[23] it inhibits GSK-3B; and it increases the ratio of neurotrophic to cell death factors. For example, it increases Bcl-2 and brain-derived neuotrophic factor (BDNF) while decreasing the apoptotic factor BAX and P-53.[25,26] Much work remains to be done to implicate or rule out these changes in the wide array of psychotrophic actions of lithium. Lithium also increases white blood cell count and platelets by increasing granulocyte-macrophage colony-stimulating factor.[30] This effect is sufficient to overcome the benign white count suppression of carbamazepine.[31,32]

The potential clinical relevance of this finding is also evidenced by the data that pretreatment with lithium can decrease the size of a cerebral infarct following middle cerebral artery ligation and decrease the amount of neurological deficit.[33] Given the new findings of altered size of crucial structures involved in emotion regulation in the affective disorders, including amygdale,[34,35] hippocampus,[36] and prefrontal cortex,[4] one can only wonder whether lithium's neuroprotective effects could alter some of these putative neuronal- or glial-based deficits in central nervous system structure and function. The preliminary evidence supports this proposition because lithium increases patients' NAA, a marker of neuronal integrity and, grey matter volume as well.[37,38]

Carbamazepine, valproate, and lamotrigine

As seen in Table 6.2.4.2, since lithium, carbamazepine, and valproate share a group of effects in common, one wonders if they are related to their global antimanic/antidepressant effects. Notable differences among these agents are also present, perhaps also related to some of the differences in therapeutic targets and comorbidities seen in Table 6.2.4.1.

Given the shared effects of carbamazepine and lamotrigine in potent sodium channel blockade and subsequent decreased release of glutamate, one wonders about mechanisms that account for their difference in the epilepsies (carbamazepine exacerbates while lamotrigine improves absence seizures) and bipolar disorder (lamotrigine is a better antidepressant than antimanic). Potential candidates are the effects on different calcium channel subtypes (N,P) and ability to inhibit GABA release, but factors critical to lamotrigine's antidepressant effects remain to be elucidated.[39]

L-type calcium channel blockers

There are several subtypes of voltage-dependent blockers of calcium influx which modulate the L-type channel, i.e. one with relatively long (L) opening times. These subtypes include the widely recognized phenylalkylamines typified by verapamil, the benzodiazepines typified by diltiazem, the diphenylpiperazines typified by flunarizine, and the 1,4-dihydropyridines typified by nifedipine, nimodipine, isradipine, amlodipine, nicardipine, and nitrendipine. Remarkably, even though these agents all potently bind to this voltage-dependent calcium channel and act to inhibit calcium influx, their biochemical and physiological effects are very different, as noted above.

Verapamil and the phenylalkylamines are charged and bind at the outer portion of the calcium channel, while the dihydropyridines, which are uncharged (except amlodipine), bind deeper inside the calcium channel. These different membrane properties and binding characteristics result in a different profile of effects for the dihydropyridines compared with the phenylalkylamines, including increased lipid solubility. It is possible that these differences relate to the increased effectiveness of the dihydropyridines on depression[40] and ultra rapid cycling seen in a small subgroup of lithium-refractory bipolar patients.

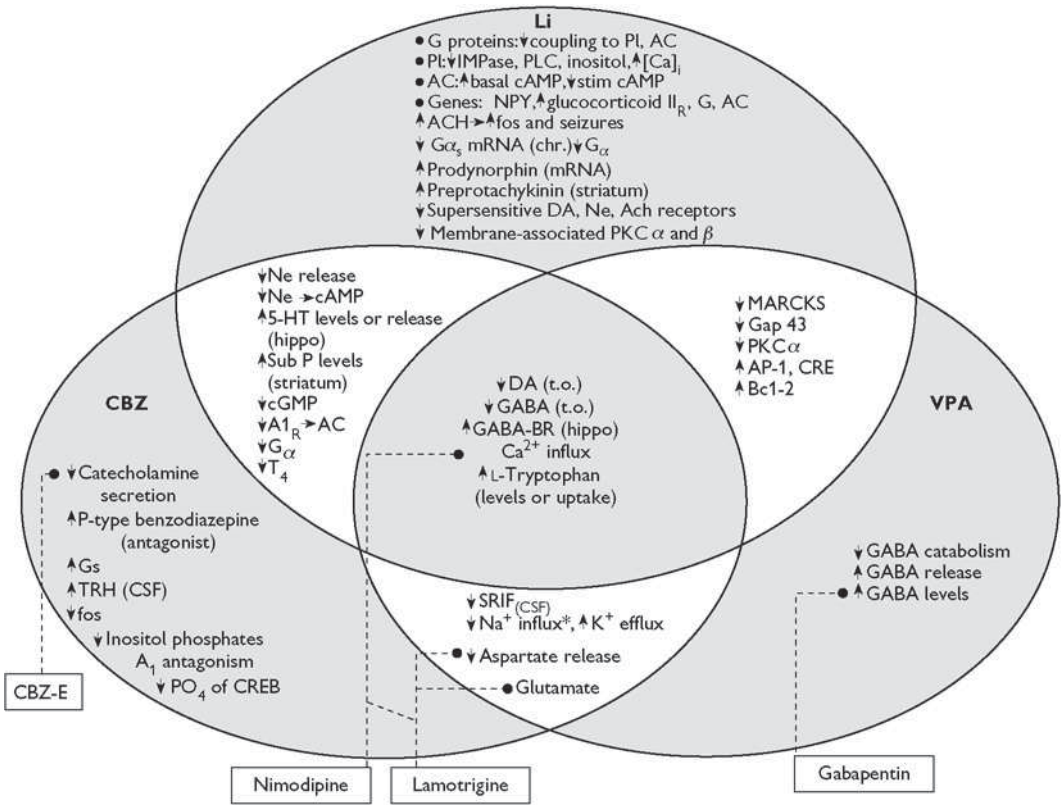


Fig. 6.2.4.1 Common and differential mechanism of mood stabilizers: PI, phosphoinositol; AC, adenylate cyclase; IMPase, inosital monophosphatase; PLC, phospholipase C; cAMP, cyclic adenosine monophosphate; NPY, neuropeptide Y; ACH, acetylcholine; $G\alpha_s$, G protein alpha (s) subunit; DA, dopaminergic; Ne, noradrenergic; PKC, protein kinase C; $A1_R$, adenosine A1 receptors; T_4, thyroxine; Gap 43, growth-associated protein 43; CRE, cyclic response element; CBZ, carbamazepine; TRH, throtrophin-releasing hormone; CREB, cyclic response element binding protein; VPA, valproate; SRIF, somatosatin; t.o., turnover.

Pharmacokinetics, dosage, and administration

Lithium

Lithium blood levels are conventionally described as being therapeutic from 0.5 to 1.2 mmol/l. Within this range, there is a wide agreement that higher doses are associated with increasing numbers of side-effects, but there is disagreement as to whether this is uniformly accompanied by a better therapeutic effect. Gelenberg *et al.*[12] reported that higher doses of lithium in the 0.8 to 1.0 mmol/l range averaging 0.83 mmol/l were 2.6 times as effective as doses achieving blood levels in the lower therapeutic range of 0.4 to 0.6, averaging 0.54 mmol/l. **This increased efficacy** (i.e. lower risk of relapse) **was achieved at the cost of a greater frequency of side effects**. Three times as many patients given doses in the high range withdrew from the study due to side effects. A series of other studies suggest that low to moderate doses of lithium may be just as effective as those in the higher range.[13]

Data of Kleindeinst *et al.*[41] further suggest a differential dose/effect relationship for lithium in mania versus depression prophylaxis. They found higher doses/blood levels more effective in preventing manic episodes while, **paradoxically, lower doses/levels were more effective in preventing depressions**. There is a considerable time lag before lithium reaches steady-state levels and some attempts at lithium loading with large doses from the outset have not been successful. A certain amount of time is needed for lithium to gain access into the central nervous system compartment and steady-state levels are not typically achieved until five half-lives or approximately six days. With the advent of magnetic resonance spectroscopy it has been found that lithium levels in the brain are half those in serum and, based on one small study, may be better correlated with the degree of clinical response than serum levels. **Older individuals have higher intracellular lithium levels**, which may account for the observations of **toxicity at apparently therapeutic blood levels**.

While lithium has traditionally been administered in two, three, or four times daily dosing regimens in attempts to achieve the most consistent and stable blood levels possible, several findings have led to changes in this pattern of dosing. Extended-release preparations are now available and suitable for twice-daily dosing. Even with the original preparations of lithium carbonate many clinicians and investigators have administered lithium in single nighttime doses in order to achieve the highest blood levels (and the potential for side effects) during sleep, and utilize the much lower trough levels at a time when the kidney, for example, can be relatively spared from continuously high lithium levels. Some investigators feel that this might be associated with lower long-term renal side effects, and preliminary data suggest that this paradigm may not be associated with any loss of clinical efficacy. Given lithium's unique anti-suicide

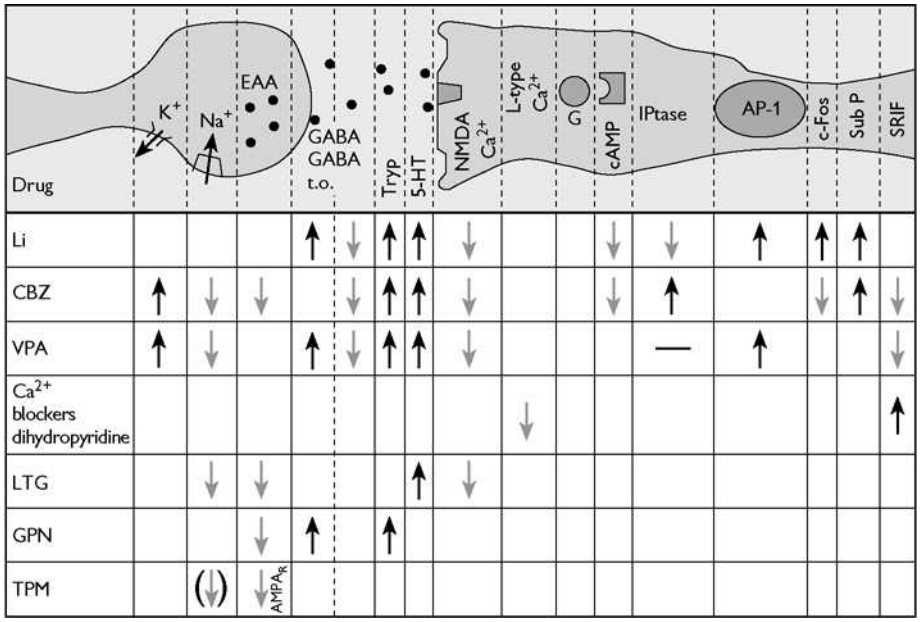


Fig. 6.2.4.2 Mechanisms of mood Stabilization. Depicted schematically at the top of the figure is a synapse with various types of channels, neurotransmitters, and proteins associated with the mechanisms of action of the mood stabilizers listed in the table below. **Row headings**: Li, lithium; CBZ, carbamazepine; VPA, valproate; LTG, lamotrigine; GPN, gabapentin; TPM, topiramate. **Column headings**; K^+ efflux; Na^+ influx; EAA, excitatory amino acids; GABA, γ-aminobutyric acid; GABA t.o., GABA turnover; Tryp, tryptophan; 5-HT, serotonin; NMDA Ca^{2+}; L-type Ca^{2+}; G, G protein; cAMP, cyclic adenosine monophosphate; IPtase, inositol phosphatase; AP-1, activator protein 1; Sub P, substance P; SRIF, somatostatin. Arrows indicate increases or decesases in substance/activity.

effects and strong evidence of neurotrophic and neuroprotective effects in animals at clinically relevant blood levels of 0.5 mmol/l, *administering lithium at whatever doses/blood levels are not associated with side effects would appear to have considerable merit*.

Carbamazepine, valproate, lamotrigine

Carbamazepine, in contrast to oxcarbazepine, is a **potent inducer of CYP 3A4** hepatic enzymes, accounting for **auto-induction after one to three weeks of treatment.**[8] With immediate release preparations, B.I.D or T.I.D. dosing is required, but with long-acting preparations such as Equetro®(Tegretol Retard®), all H.S. dosing should be considered. The dose and blood level to side-effects relationships are highly variable and individual. In non-emergency situations, 200 mg H.S. may be considered as a starting dose with slow upward titration to avoid or accommodate to side effects. In manic inpatients initial dosing of 600 mg to 800 mg on day one may be tolerated.

While blood levels of 4–12 μg/ml are touted as therapeutic in epilepsy, there is no relation of blood level to degree of clinical response across either seizure disorder or affectively ill patients. Therefore, **individualized titration to good clinical effect and minimal side-effects** burden is more appropriate than compulsive blood level monitoring.

Valproate may be given in loading doses of 10 to 15 mg/kg. Efficacy is usually observed between 50–120 mg/ml with a target of 80 mg/ml or more in mania. *Single nighttime doses can be employed* with both immediate release and extended release preparations, which may be better tolerated.

Lamotrigine must be dosed and titrated very slowly in an attempt to avoid the occurrence of a **serious rash**. A typical procedure is to start with 25 mg/day for two weeks and then 50 mg/day for two weeks, and then increase by 25 mg (preferably) to 50 mg/week thereafter. A target dose is about 200 mg/day, but increases to 400 mg/day in those showing partial responses is often tolerated. The rate of titration and target dose of lamotrigine should be halved on **valproate** (which **doubles lamotrigine levels**) and can be doubled on carbamazepine and related potent enzyme inducers.

L-type calcium channel blockers

The pharmacokinetics of the dihydropyridine L-type calcium channel blockers differ markedly. Nimodipine has a T½ of 1–2 h requiring T.I.D. dosing to peak total doses of 240 to 480 mg/day. Isradipine's T½ is 8 h, allowing B.I.D. dosing to a peak dose of 15 mg/day. Notably, amlodipine has a longer half-life and is suitable for single nighttime or twice-daily dosing in comparison with the better psychiatrically studied drugs nimodipine and isradipine. However, as long-acting preparations of these compounds become available, the importance of this half-life dissociation among the different compounds may dissipate. In our patient cohort at the NIMH, the pharmacokinetics of nimodipine (Bay e 9736) and its several metabolites (Bay o 1762, Bay m 5397, and Bay m 8922) were characterized by a rapid peaking (in 30–45 min) and decline in the second hour in response to a 60 mg challenge dose at steady-state blood levels after an overnight fast. Additional small secondary peaks occurred when usual dosing of four-times-a-day was resumed. There were no notable differences between capsule and tablet preparations. The phenylalkylamine verapamil has a 5–8 h half-life, also requiring T.I.D. dosing to a target peak daily dose of 480 mg/day.

Side effects

Lithium

Since lithium has been in use for much of the latter half of the twentieth century, its side-effects profile has been well described. Tremor and gastrointestinal distress, particularly diarrhoea, are generally dose-related, but some patients can have idiosyncratic sensitivity to these side effects, even at relatively low doses. Lithium-induced tremor can be countered with the beta-blocker propranolol in doses of 10 mg four times/day.

Side effects most likely to be associated with non-compliance or discontinuation of the drug include a sense of psychomotor slowing, cognitive dulling, acne or psoriasis, and weight gain. There is preliminary evidence that the anticonvulsants topiramate or zonisamide may help to reverse or stabilize lithium-related weight gain. Cognitive dulling could be treated with dose lowering, assessment of thyroid function, and T_3 (25–50 μg) augmentation even with normal thyroid indices, folate augmentation, and, potentially, an acetylcholinesterase inhibitor such as donepezil (Aricept®).

Lithium interferes with the actions of ADH (i.e. vasopressin) because of its ability to block vasopressin-induced adenylate cyclase. A syndrome of **reversible diabetes insipidus** is thus induced which, in most patients, is not problematic, although in a small percentage of patients, excretion of large volumes of urine can be extreme, inconvenient, and disruptive of normal social routines and sleep. This can be countered with amiloride or the thiazide diuretics (however, the latter also increase lithium levels).

Lithium is clearly able to induce thyroid dysfunction with increases in thyroid-stimulating hormone, sometimes proceeding to more full-blown evidence of chemical hypothyroidism. The threshold for treating lithium-related increases in thyroid-stimulating hormone has not been definitively identified, but with some evidence of lower levels of free thyroxine being associated with increased levels of depression and other low thyroid indices being associated with increased cognitive dysfunction, **replacement of thyroid hormones would appear indicated as thyroid-stimulating hormone begins to exceed normal levels**. Whether thyroid supplementation can reverse or prevent these lithium-related abnormalities remains to be directly assessed in prospective studies.

Some investigators suggest that long-term lithium may be associated with **slowly increasing creatinine levels** and a decrease in creatinine clearance.[40] The incidence of these glomerular filtration abnormalities in lithium-treated patients compared with age- and gender-matched controls remains controversial, as does the mode of treatment in the face of progressive changes in these indices. Given the availability of other potential mood-stabilizing agents, a reduction in lithium levels and supplementation or switching to other agents would be a conservative measure, if tolerated. However, others recommend careful monitoring of continued lithium therapy because the effects of lithium discontinuation on creatinine levels are highly inconsistent.

Severe episodes of lithium intoxication are to be avoided since they can be associated with a syndrome of irreversible cerebellar dysfunction.[42] The use of lithium with very high dose neuroleptic treatment is also to be avoided since occasional idiosyncratic and irreversible organic brain syndromes have resulted on rare occasions. Marked EEG changes and tonic-clonic seizures were also observed with the combination of lithium and clozapine. Lithium can be associated with alterations in calcium homeostasis and frant hyperparathyroidism. As mentioned above, lithium can increase white cells and platelets via its action on granulocyte-macrophage colony stimulating factor.

Valproate has many lithium-like side effects which may also be additive in combination treatment. **Valproate has a black box warning for rare hepatitis/pancreatitis and should not be given to children below two years of age**. It can also cause low platelets, and signs of bleeding tendency should be attended to. Because **valproate increases homocysteine, routine supplementation with folate (and possibly also B6 and B12 in women of child bearing age) would appear prudent**. Valproate can cause asymptomatic to symptomatic hyperammoniumemis; treatment with l-carnitine may be helpful. Zinc and selenium are anecdotally touted as preventing alopecia, but systematic evidence is lacking. Data are mixed as to whether valproate increases testosterone and causes the polycystic ovary syndrome (PCOS). Hirsutism is rarely seen in affectively ill patients and birth control pills will prevent PCOS. Valproate does increase the already high rates of menstrual irregularities seen in women with bipolar disorder.

Carbamazepine has less weight-gain liability than lithium or valproate, but has greater potential for rash and rare but serious hematological problems. **Carbamazepine** routinely causes a benign **drop in white blood cells via its effects on colony stimulating factor; this can be countered by lithium**. More problematic are agranulocytosis and aplastic anemia, estimated to occur in 1 in 20 000 to 50 000 patients. **Patients should be warned to consult their physician** if they develop a fever, sore throat, or other infection that could emanate from a low white blood count, or bleeding gums or petechiae that could reflect low platelets. **Hyponatremia is more common with oxcarbazepine** than carbamazepine, but the hyponatremia of carbamazepine can be treated or prevented with lithium or demeclocycline. Low T_4 on carbamazepine is usually not reflective of hypothyroidism because TSH is not increased, BMR is not decreased, and the degree of drop in T_4 and free T_4 may even be correlated with degree of antidepressive effect.

The **major side-effects concern of lamotrigine is that of a severe rash**, estimated to occur in 1 of 5 000 adults and 1 of 2 500 children. Even a benign rash should lead to drug discontinuation because there is no way to predict when a rash may progress to a Stevens-Johnson syndrome or Toxic Epidermal Necrolysis. Otherwise, the side-effects profile of lamotrigine fits well with bipolar depression treatment because the drug is weight-neutral, non-sedating, and without sexual dysfunction or endocrine dysregulation.

L-type calcium channel blockers

The side-effects profile of nimodipine and related dihydropyridine L-type calcium channel blockers differs considerably from that of lithium. These drugs are primarily used in cardiology for their anti-hypertensive and anti-arrhythmogenic effects. As such, they may cause **side effects related to hypotension including dizziness and tightness in the chest**. Unlike lithium, they are not typically associated with gastrointestinal distress and tend to be slightly constipating rather than associated with diarrhea. Therefore, they might be able to replace some component of lithium's therapeutic action, as noted below, potentially without exacerbating some of lithium's related side effects. Although these agents are often used in migraine prophylaxis, they can also be associated with headache on rare occasions. Redness and erythema with excessive warmth in

the pretibial areas is also an occasional side effect of these agents, as is, more rarely, edema itself.

Nimodipine and related dihydropyridine L-type calcium channel blockers do not appear to share lithium's ability to induce cognitive slowing and, in fact, these agents have been reported to improve performance in some preclinical models of learning and memory deficits as well as in some clinical studies of patients with Alzheimer's disease. This might be related to nimodipine's ability to increase somatostatin upon chronic administration,[43] although other mechanisms remain to be explored.

Also, in contrast to lithium, which is associated with a small incidence of Epstein's anomaly, verapamil is not teratogenic. While systematic data are not available for the other dihydropyridine L-type calcium channel blockers, it is hoped that they will prove to be as safe as verapamil appears to be.

Indications and contraindications

Lithium

The double-blind controlled studies of lithium in acute mania were positive several decades ago, and in the largest randomized study comparing lithium with placebo in a study primarily designed to evaluate the acute antimanic efficacy of valproate, lithium and valproate appeared to show approximately equal efficacy and both were superior to placebo.[44] However, there appear to be a number of subtypes of illness with consistently higher or lower rates of response. As summarized in Table 6.2.4.1, a useful rule of thumb is that lithium is relatively less effective, with a low effectiveness rate of around 30 per cent in acute manic syndrome characterized by anxiety and dysphoria, comorbid substance abuse, comorbid medical conditions, the pattern of illness of depression-mania-well intervals as opposed to mania-depression-well intervals, in those with a negative family history of bipolar illness in first-degree relatives, in those with evidence of **EEG and neurological dysfunction, and in those with a pattern of rapid cycling or multiple prior episodes.**[2]

Lithium used to be considered contraindicated in pregnancy, but with the recognition that major cardiac malformations such as Epstein's anomaly are rare (1 in 1 200 births), many clinicians and patients are deciding to continue lithium when it appears important to continued mood stability.

L-type calcium channel blockers

As discussed earlier, the evidence of verapamil's acute antimanic efficacy is derived from a considerable series of small double-blind studies. While the initial studies were unequivocally positive, more recent studies have not been as positive comparing either verapamil with placebo or verapamil with lithium. Only very preliminary evidence is available for the acute antimanic efficacy of nimodipine, with the open study of Brunet *et al.*[45] being positive in six of six individuals.[18,63] In placebo-controlled studies using an off-on-off-on design, 10 of 30 patients with refractory recurrent affective disorder responded and, in may instances, the antimanic efficacy was demonstrated both in the on-phase and with symptom exacerbation in the off-phase.[46] However, many of these individuals had ultra-rapid or ultra-ultra-rapid (ultradian) cycling patterns and most of the efficacy data reflected effective pharmacoprophylaxis of mania rather than acute antimanic efficacy *per se*.

Dubovsky[47] observed that a prior history of lithium response appears to be associated with a good response to verapamil. However, with the dihydropyridines we have observed some instances of response in those who were previously non-responders to lithium.[46] Among those responsive to the drug were patients with ultra-rapid and ultradian cycling and those with a pattern of recurrent brief depression. As indicated above, a number of these individuals were rechallenged and **responsiveness was confirmed in an off-on-off-on design**. Whether those bipolar patients with increased intracellular calcium would be among those responsive to the L-type calcium channel blockers also awaits completion of such a study, which is now under way.

A promising area of investigation is work using brain imaging, which has found that depressed patients with the classical pattern of relative frontal hypometabolism, and especially in the left insula, were among those who responded best to nimodipine, while equally depressed patients with left insula hypermetabolism responded to carbamazepine.[48] These data raise the possibility that regional topographies of blood flow or metabolism might ultimately help identify a subgroup of patients more responsive to the calcium-channel blocking agents. Also potentially helpful are data showing that nimodipine increases somatostatin in the CSF, and in one small study, that those with lower baseline CSF somatostatin were better responders to the drug.[49]

Interactions

Lithium

Owing to its renal excretion, lithium has renally-mediated rather than hepatically-mediated drug-drug interactions. Lithium excretion is decreased by medications such as thiazides, non-steroidal anti-inflammatory drugs, angiotensin-converting enzyme inhibitors, and, to a lesser extent, furosemide, and by physiological states such as dehydration, advanced age, and renal disease.[50] Owing to lithium's poor therapeutic index, these interactions can result in clinical lithium toxicity unless appropriate dosage adjustment is made. In contrast, lithium clearance is less consistently affected by amiloride, aspirin, and sulindac, and increased with other medications with diuretic effects such as acetazolamide, mannitol, aminophylline, caffeine, and theophylline, as well as during pregnancy.

Valproate, carbamazepine, lamotrigine

Valproate is FDA-approved for migraine prevention and treatment of acute mania. However, it is **widely considered a mood stabilizer and used in long-term prophylaxis of both mood phases**. A modicum of data support acute and prophylactic efficacy in depression. **Valproate decreases alcohol intake** in bipolar patients[51] with this comorbidity and is useful in a wide range of anxiety syndromes.

Valproate is contraindicated in pregnancy because of a several percentage risk of neural tube defects (spina bifida) and, more recently noted, a substantial (20 per cent) incidence of other adverse events,[52] as well as moderate to severe developmental delay in a sizeable number of children.

Carbamazepine is approved for trigeminal neuralgia and the long-acting preparation of carbamazepine (Equetro®/Tegretol Retard®) is approved only in acute mania, but can be considered in long-term prophylaxis for those not responsive to lithium or valproate.

Its profile of effectiveness is almost the converse of that of lithium, with better effects in those with bipolar II illness, anxiety, and substance abuse comorbidity, mood incongruent delusions,[53] and a negative family history of bipolar disorder in first-degree relatives. Carbamazepine should not be used in pregnancy because it carries about a 0.5 per cent risk of neural tube defects.

Lamotrigine is unique because it is approved for only **prevention of depressed episodes** and to a lesser extent, manic and mixed episodes. Robust efficacy in acute depression has been seen in three studies, but not in four others, although the meta analysis remains significantly positive. A small series of patients studied in Canada identified correlates of positive response, including continuous cycling pattern and a **personal and positive family history of anxiety disorders**.[54,55] Only one of five pregnancy case registries showed a significance for the increased occurrence of cleft lip and palate, and this agent clearly appears safer than valproate or carbamazepine. Its overall serious adverse event percentage of 1 per cent does not appear to be different from that of the general population.

L-type calcium channel blockers

Different calcium-channel blockers differentially affect carbamazepine levels. While verapamil and diltiazem increase carbamazepine levels substantially, potentially causing toxicity, this is not the case with nimodipine, isradipine, or amlodipine. Preliminary data from our group and others suggest that carbamazepine decreases nimodipine levels after a 60 mg challenge dose. In our study, group mean peak nimodipine levels during treatment with carbamazepine were about one-half those observed during treatment with nimodipine alone, although this finding showed only a trend level of significance, probably due to small sample size.

When calcium channel blockers are added to ß-adrenergic blocking agents, depression of ventricular function, cardiac slowing, and atrioventricular block can result. Combining calcium channel blockers with ß-adrenergic blocking agents may produce hypotension. Verapamil and nitrendipine increase plasma concentrations of digoxin and produce bradycardia, hypotension, or atrioventricular block.

Effects of withdrawal

Lithium

In addition to the variety of predictors of relative lithium nonresponsiveness from the outset, two relatively new and different mechanisms for the development of treatment resistance or loss of efficacy have been uncovered during long-term follow up in patients who are initially responsive. The first of these is the apparent development of tolerance characterized by an increasing frequency and/or severity of breakthrough episodes despite good compliance and consistent maintenance of lithium blood levels. In a group of 66 patients referred to the NIMH because of lithium nonresponsiveness, 23 patients **(34.8 per cent) displayed this apparent tolerance pattern**.[55] Although it has not been systematically studied, the initial therapeutic manoeuvres in the face of such loss of efficacy at maximum tolerated doses would appear to be augmenting lithium's effects with other putative mood-stabilizing agents with different mechanisms of action, and if lithium should be discontinued, a consideration of its reinstitution with a hope for renewal of responsivity.

In contrast with this tolerance pattern in which patients suffer breakthrough episodes despite remaining under treatment, the phenomenon of lithium-discontinuation-related refractoriness refers to a small group of patients who have done extremely well on their long-term lithium, discontinue the drug, suffer additional relapses, and then fail to re-respond once lithium is reinstituted. This phenomenon accounted for nine of the 66 patients (13.6 per cent) who presented to us as lithium-refractory.[56] The average time well on lithium was 6.6 years, substantially greater than the average well interval of 1.5 years prior to instituting lithium therapy, strongly suggesting that lithium had been effective in these individuals, and if they had remained on the drug, they might have remained well. Sadly, for each of these individuals, this did not prove to be the case.

One patient had been well on lithium for more than 16 years and tapered lithium slowly, suggesting that neither the duration of time well nor the use of a slow taper would necessarily prevent the development of discontinuation-induced refractoriness. A number of other investigative groups have observed such a phenomenon.[57] In these studies, discontinuation-induced refractoriness occurred in anywhere from 3.6 per cent to 18.6 per cent of patients, with a total of 39 of 321 (12.1 per cent), and 12 of 92 (13 per cent) in studies that tracked responders only. Even if it only occurs in about 10 per cent of patients who discontinue their lithium, it would nevertheless appear to be of considerable clinical import and should be included in the informed consent process so that the patient has all of the available data when making decisions of whether or not to continue treatment. That is, in considering the risk-benefit of stopping lithium, **a patient should not only know the very high risk of relapse** (50 per cent in the first five months after discontinuing lithium, 80–90 per cent after 1.5 years,[58] but also that there is **no guarantee that responsivity would be as rapid, robust, or complete** as previously experienced, and that a small subgroup of individuals, perhaps as many as 10 per cent, will not achieve the same good response that they had previously.

Valproate, carbamazepine, lamotrigine

Valproate will displace carbamazepine from its protein binding and inhibit the epoxide hydrolase from converting the active epoxide to the inactive diole. Thus, **carbamazepine dose will need to be decreased when used in combination with valproate**. **Valproate will double lamotrigine levels** such that dosing should be one-half that of normal.

Carbamazepine as a potent 3A4 inducer has many drug-drug interactions. It will decrease levels of oestrogen such that **high dosage forms need to be used with birth control pills**. It will lower levels of lamotrigine, haloperidol, aripiprazole, and many other compounds.

Inhibition of 3A4 will notably increase carbamazepine levels and potentially cause toxicity if a patient is at or near their side-effects threshold. **Erythromycin** and its analogues, **verapamil, and diltiazem** are such examples, and patients should be warned to check with their pharmacists about its many other interactions. Side effects can be avoided with lower carbamazepine doses in advance of such drugs being administered.

Each of these drugs can be discontinued rapidly without a major withdrawal syndrome. Tolerance to the efficacy of each of these

agents has been observed in isolated cases or small series. Increasing doses and/or switching to or adding other agents without cross-tolerance (mechanistically different) are typically used clinically [8] In instances of tolerance development a period of time off the new ineffective drug may be associated with transient renewal of efficacy. Theoretical, but unproven ways of slowing or minimizing tolerance development include: using more maximally tolerated doses, rather than minimally effective ones; holding doses stable and not decreasing them unnecessarily; using combination of drugs; and treating earlier as opposed to later in the course of illness.

L-type calcium channel blockers

Because so few patients have been studied with the L-type calcium channel blockers in long-term prophylaxis, it is uncertain to what extent patients may become tolerant to these agents. However, since tolerance has been observed to virtually every other putative mood-stabilizing agent, it is likely that this will also occur with nimodipine and related agents. Similarly, it is uncertain whether the phenomenon of discontinuation-related refractoriness observed with lithium would extend to the class of L-type calcium channel blockers; while we have not observed this phenomenon in our cohort of nimodipine responders (63), only a small group of patients has been studied to date.

Further information

Post, R.M., and Leverich, G.S. (2008). *Treatment of Bipolar Illness: A Case Book for Clinicians and Patients*. WW Norton, Inc. (in press).

Post, R.M., Altshuler, L.L. (2005). Mood disorders: treatment of bipolar disorders. In: *Comprehensive Textbook of Psychiatry*. (eds. B.J. Sadock and V.A. Sadock), pp. 1661–707. Lippincott Williams & Williams, New York.

Post, R.M., Speer, A., Leverich, G.S. (2006).Complex combination therapy: the evolution toward rational polypharmacy in lithium-resistant bipolar illness. In *Bipolar Psychopharmacotherapy: Caring for the Patient*. (eds. H. Akiskal and M. Tohen), pp. 135–67. John Wiley & Sons.

References

1. Post, R.M., and Leverich, G.S. (2007). *Treatment of Bipolar Illness: A Case Book for Clinicians and Patients*. WW Norton, Inc. (in press).
2. Post, R.M., Speer, A., Leverich, G.S., *et al.* (2006).Complex combination therapy: the evolution toward rational polypharmacy in lithium-resistant bipolar illness. In *Bipolar Psychopharmacotherapy: Caring for the Patient*. (eds. H. Akiskal and M. Tohen), pp. 135–67. John Wiley & Sons.
3. Post, R.M. and Luckenbaugh, D.A. (2003). Unique design issues in clinical trials of patients with bipolar affective disorder. *Journal of Psychiatric Research*, **37**(1), 61–73.
4. Post, R.M., Leverich, G.S., Weiss, S.R., *et al.* (2003). Psychosocial stressors as predisposing factors to affective illness and PTSD: potential neurobiological mechanisms and theoretical implications. In: *Neurodevelopmental Mechanisms in Psychopathology*. (eds. D. Cicchetti and E. Wilker). pp. 491–525. Cambridge University Press, New York.
5. Post, R.M. (2007a). Kindling and sensitization as models for affective episode recurrence, cyclicity, and tolerance phenomena. *Neuroscience and Biobehavioral Reviews*, **31**(6), 858–73.
6. Leverich, G.S., Post, R.M., Keck, P.E. Jr., *et al.* (2007). The poor prognosis of childhood-onset bipolar disorder. *Journal of Pediatrics*,**150**(5), 485–90.
7. Post, R.M. (2007b). Role of BDNF in bipolar and unipolar disorder: clinical and theoretical implications. *Journal of Psychiatric Research*, **41**, 979–90.
8. Post, R.M., Ketter, T.A., Uhde, T., *et al.* (2007). Thirty years of clinical experience with carbamazepine in the treatment of bipolar illness: principles and practice. *CNS Drugs*, **21**(1), 47–71.
9. Post, R.M. (2004). The status of the sensitization/kindling hypothesis of bipolar disorder. *Current Psychosis and Therapeutics Reports*, **2**, 135–41.
10. Kessing, L.V., and Andersen, P.K. (2004). Does the risk of developing dementia increase with the number of episodes in patients with depressive disorder and in patients with bipolar disorder? *Journal of Neurology, Neurosurgery, and Psychiatry*, **75**(12),1662–66.
11. Schou, M. (1997). Forty years of lithium treatment. *Archives of General Psychiatry*, **54**, 9–15.
12. Gelenberg, A.J., Kane, J.M., Keller, M.B., *et al.* (1989). Comparison of standard and low serum levels of lithium for maintenance treatment of bipolar disorder. *New England Journal of Medicine*, **321**, 1489–93.
13. Vestergaard, P., Licht, R.W., Brodersen, A., *et al.* (1998). Outcome of lithium prophylaxis: a prospective follow-up of affective disorder patients assigned to high and low serum lithium levels. *Acta Psychiatrica Scandinavica*, **98**, 310–15.
14. Gitlin, M.J., and Altshuler, L.L. (1997). Unanswered questions, unknown future for one of our oldest medications. *Archive General of Psychiatry*, **54**(1), 21–3.
15. Dubovsky, S.L., Murphy, J., Thomas, M. *et al.* (1992). Abnormal intracellular calcium ion concentration in platelets and lymphocytes of bipolar patients. *American Journal of Psychiatry*, **149**, 118–20.
16. Dubovsky, S.L., Franks, R.D., Allen, S. *et al.* (1986). Calcium antagonists in mania: a double-blind study of verapamil. *Psychiatry Research*, **18**, 309–20.
17. Janicak, P.G., Sharma, R.P., Pandey, G., *et al.* (1998). Verapamil for the treatment of acute mania: a double-blind, placebo controlled trial. *American Journal of Psychiatry*, **155**, 972–3.
18. Walton, S.A., Berk, M., and Brook S. (1996). Superiority of lithium over verapamil in mania: a randomized, controlled, single-blind trial. *Journal of Clinical Psychiatry*, **57**, 543–6.
19. Hoschl, C. and Kozeny, J. (1989). Verapamil in affective disorders: a controlled, double-blind study. *Biological Psychiatry*, **25**, 128–40.
20. Pazzaglia, P.J., Post, R.M., Ketter T.A., *et al.* (1993). Preliminary controlled trial of nimodipine in ultra-rapid cycling affective dysregulation. *Psychiatry Research*, **49**, 257–72.
21. Post, R.M., Pazzaglia, P.J., Ketter, T.A., *et al.* (2000). Carbamazepine and nimodipine in affective illness: effiicacy, mechanisms of action, and interactions. In *Pharmacotherapy for mood, anxiety, and cognitive disorders* (eds. S. Montgomery and U. Halbreich), American Psychiatric Press, Washington, DC.
22. Triggle, DJ. (2007). Calcium Channel Antagonists: Clinical uses-past, present, and future. *Biochemical Pharmacology*, **74,** 1–9.
23. Belmaker, R.H., Avissar, S., and Schreiber, G. (1991). Effect of lithium on human neurotransmitter receptor systems and G proteins. In *Lithium and the cell: pharmacology and biochemistry* (ed. N.J. Birch), pp. 113–19. Academic Press, London.
24. Manji, H.K., Potter, W.Z., and Lenox, R.H. (1995). Signal transduction pathways. Molecular targets for lithium's actions. *Archives of General Psychiatry*, **52**, 531–43.
25. Bebchuk, J.M., Arfken, C.L., Dolan-Manji, S., *et al.* (2000). A preliminary investigation of a protein kinase C inhibitor in the treatment of acute mania. *Archives of General Psychiatry*, **57**, 95–7.
26. Nonaka, S., Hough, C.J., and Chuang, D.M. (1998). Chronic lithium treatment robustly protects neurons in the central nervous system against excitotoxicity by inhibiting *N*-methyl-*D*-aspartate receptormediated calcium influx. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 2642–7.
27. Chen G., Zeng, W.Z., Yuan, P.X., *et al.* (1998). The mood-stabilizing agents lithium and valproate dramatically increase the levels of the neuroprotective protein BCL-2 in the CNS. *Journal of Neurochemistry*, **72**, 879–82.

28. Zarate, C.A. Jr., Singh, J.B., Carlson, P.J., *et al.* (2007). Efficacy of a protein kinase C inhibitor (tamoxifen) in the treatment of acutemania: a pilot study. *Bipolar Disorders*, **9**(6), 561–70. 29. Einat, H., Yuan, P., Szabo, S.T., *et al.* (2007). Protein kinase C inhibition by tamoxifen antagonizes manic-like behavior in rats: implications for the development of novel therapeutics for bipolar disorder. *Neuropsychobiology*, **55**(3–4), 123–31.
30. Gallicchio, V.S., Chen, M.G., and Watts, T.D. (1984). Specificity of lithium (Li+) to enhance the production of colony stimulating factor (GM-CSF) from mitogen-stimulated lymphocytes *in vitro. Cellular Immunology*, **85**, 58–66.
31. Joffe, R.T., Post, R.M., Roy-Byrne, P.P., *et al.* (1985). Hematological effects of carbamazepine in patients with affective illness. *American Journal of Psychiatry*, **142**, 1196–9.
32. Kramlinger, K.G. and Post, R.M. (1990). Addition of lithium carbonate to carbamazepine: hematological and thyroid effects. *American Journal of Psychiatry*, **147**, 615–20.
33. Chuang, D.M., Chen, R.W., Chalecka-Franaszek, E., *et al.* (2002). Neuroprotective effects of lithium in cultured cells and animal models of diseases. *Bipolar Disorders,* **4**(2):129–36.
34. Wignall, E.L., Dickson, J.M., Vaughan, P., *et al.* (2004). Smaller hippocampal volume in patients with recent-onset posttraumatic stress disorder. *Biological Psychiatry*, **56**(11), 832–6.
35. Blumberg, H.P., Fredericks, C., Wang, F., *et al.* (2005). Preliminary evidence for persistent abnormalities in amygdala volumes in adolescents and young adults with bipolar disorder. *Bipolar Disorders*, **7**(6), 570–6.
36. Sheline, YI., Sanghavi, M., Mintun, M.A., *et al.* (1999). Depression duration but not age predicts hippocampal volume loss in medically healthy women with recurrent major depression. *Journal of Neuroscience*, **19**(12), 5034–43.
37. Moore, G.J., Bebchuk, J.M., Wilds, I.B., *et al.* (2000). Lithium-induced increase in human brain grey matter. *Lancet*, **356**(9237): 1241–2.
38. Bearden, C.E., Thompson, P.M., Dalwani, M., *et al.* (2007). Greater cortical gray matter density in lithium-treated patients with bipolar disorder. *Biological Psychiatry*, **62**(1), 7–16.
39. Ketter, T.A., Wang, P.W., Becker, O.V., *et al.* (2003). The diverse roles of anticonvulsants in bipolar disorders. *Annals of Clinical Psychiatry*, **15**(2), 95–108.
40. Taragano, F.E., Bagnatti, P., Allergri, F., *et al.* (2005). A double-blind, randomized clinical trial to assess the augmentation with nimodipine of antidepressant therapy in the treatment of 'vascular depression'. *International Psychogeriatrics*, **17**, 487–498.
41. Kleindienst, N., Engel, R., Greil, W., *et al.* (2005). Which clinical factors predict response to prophylactic lithium? A systematic review for bipolar disorders. *Bipolar Disorders*, **7**(5), 404–17.
42. Kores, B. and Lader, M.H. (1997). Irreversible lithium neurotoxicity: an overview. *Clinical Neuropharmacology*, **20**, 283–99.
43. Pazzaglia, P.J., George, M.S., Post, R.M., *et al.* (1995). Nimodipine increases CSF somatostatin in affectively ill patients. *Neuropsychopharmacology*, **13**, 75–83.
44. Bowden, C.L., Brugger, A.M., Swann, A.C., *et al.* (1994). Efficacy of divalproex vs. lithium and placebo in the treatment of mania. The Depakote Mania Study Group. *Journal of the American Medical Association*, **271**, 918–24.
45. Brunet, G., Cerlich, B., Robert, P., *et al.* (1990). Open trial of a calcium antagonist, nimodipine, in acute mania. *Clinical Neuropharmacology*, **13**, 224–8.
46. Pazzaglia, P.J., Post, R.M., Ketter, T.A., *et al.* (1998). Nimodipine monotherapy and carbamazepine augmentation in patients with refractory recurrent affective illness. *Journal of Clinical Psychopharmacology*, **18**, 404–13.
47. Dubovsky, S.L. (1995). Calcium channel antagonists as novel agents for manic-depressive disorder. In *Textbook of psychopharmacology* (eds. A.E. Schatzberg and C.B. Nemeroff), pp. 377–88. American Psychiatric Press, Washington, DC.
48. Ketter, T.A., Kimbrell, T.A., George, M.S., *et al.* (1999). Baseline cerebral hypermetabolism associated with carbamazepine response and hypometabolism with nimodipine response in mood disorders. *Biological Psychiatry*, **46**, 1364–74.
49. Frye, M.A., Pazzaglia, P.J., George, M.S., *et al.* (2003). Low CSFsomatostatin associated with response to nimodipine in patents with affective illness. *Biological Psychiatry*, **53**(2), 180–3.
50. Muller-Oerlinghausen, B. (1999). Drug interactions with lithium: aguide for clinicians. *CNS Drugs*, **11**, 41–8.
51. Salloum, I.M., Cornelius, J.R., Daley, D.C., *et al.* (2005). Efficacy of valproate maintenance in patients with bipolar disorder and alcoholism: a double-blind placebo-controlled study. *Archives of General Psychiatry*, **62**(1), 37–45.
52. Meador, K.J., Baker, G.A., Finnell, R.H., *et al.* (2006). In utero antiepileptic drug exposure: fetal death and malformations. *Neurology*, **67**(3), 407–12.
53. Greil, W., Kleindienst, N., Erazo, N., *et al.* (1998). Differential response to lithium and carbamazepine in the prophylaxis of bipolar disorder. *Journal of Clinical Psychopharmacology*, **18**(6), 455–60.
54. Passmore, M.J., Garnham, J., Duffy, A., *et al.* (2003). Phenotypic spectra of bipolar disorder in responders to lithium versus lamotrigine. *Bipolar Disorders*,**5**(2), 110–4.
55. Grof, P. (2003). Selecting effective long-term treatment for bipolar patients: monotherapy and combinations. *Journal of Clinical Psychiatry*, **64** Suppl 5, 53–61.
56. Post, R.M., Leverich G.S., Altshuler, L., *et al.* (1992). Lithium-discontinuation-induced refractoriness: preliminary observations. *American Journal of Psychiatry*, **149**, 1727–9.
57. Koukopoulos, A., Reginaldi, D., Minnai, G., *et al.* (1995). The long term prophylaxis of affective disorders. *Advances in Biochemical Psychopharmacology*, Gessa, Fratta, Pani & Serra (eds)., *Depression and Mania: From Neurobiology to Treatment*, Vol 49.
58. Suppes, T., Baldessarini, R.J., Faedda, G.L., *et al.* (1991). Risk of recurrence following discontinuation of lithium treatment in bipolar disorder. *Archives of General Psychiatry*, **48**, 1082–4.

6.2.5 Antipsychotic and anticholinergic drugs

Herbert Y. Meltzer and William V. Bobo

Introduction

The discovery by Delay and Denicker in 1953 that chlorpromazine was highly effective in alleviating delusions, hallucinations, and disorganized thinking, was the seminal breakthrough in the treatment of schizophrenia, the first agent to produce sufficient relief of core psychotic symptoms to permit life outside of institutions for many patients with schizophrenia, and even a return to a semblance of function within normal limits. Chlorpromazine and the other related typical antipsychotic drugs which were introduced over the next 30 years have proven to be of immense benefit to vast numbers of people who experience psychotic symptoms as a

component of a diverse group of neuropsychiatric and medical disorders, as well as drug-induced psychoses. These drugs have been invaluable in providing clues to the aetiology of schizophrenia and other forms of mental illness with psychotic features and as tools in understanding fundamental neural processes, especially those involving dopamine, a key neurotransmitter involved in psychosis. This class of drugs has now been supplanted by the so-called atypical antipsychotic drugs, of which clozapine is the prototype. This chapter will describe the various classes of antipsychotic agents, with emphasis on the atypical antipsychotic drugs, their benefits and adverse effects, recommendations for use in clinical practice, and mechanism of action. The drugs used to treat the extrapyramidal side-effects (EPS) produced mainly by the typical antipsychotic drugs are also considered.

The classes of antipsychotic drugs

Antipsychotic drugs have been classified into two broad categories: typical and atypical.[1] Typical antipsychotic drugs are those which (typically) produce EPS at clinically effective doses, including parkinsonism (muscle rigidity, tremor, bradykinesia), acute dystonic reactions, dyskinesias, akathisia (restlessness), and tardive dyskinesia. They are also called neuroleptics because of their inhibitory effect upon locomotion activity. They are sometimes referred to as first generation antipsychotic drugs, but this has multiple problems as a class designation. The prototype of the atyipcal class of agents is clozapine which was first discovered during the early stages of the development of the drugs called first generation agents. The major mode of action of typical neuroleptics is to block dopamine D_2 receptors in the limbic system, which includes the nucleus accumbens, stria terminalis, and amygdala.

The typical antipsychotic drugs are members of a variety of chemical families (Table 6.2.5.1). They vary in affinity for the D_2 receptor, with low affinity drugs such as chlorpromazine, which require high doses for clinical efficacy, to high affinity drugs such as haloperidol, which are effective at lower doses (Table 6.2.5.1). Kapur and Seeman[3] have proposed that the rate of dissociation of all antipsychotic drugs from the D_2 receptor provides the basis for the distinction between typical and atypical antipsychotic drugs, with atypical antipsychotic drugs dissociating more rapidly. While this is true for clozapine and quetiapine, the atypical drugs risperidone, sertindole, olanzapine and asenapine dissociate no more rapidly or even slower than haloperidol. As such, 'fast dissociation' cannot provide the pharmacological basis for atypicality for most of the drugs that are considered atypical.

Low-potency typical neuroleptic agents are those in which the usual dose range in schizophrenia is equal to or greater than 200 mg/day, while mid- to high-potency agents are those in which the dose range is between 2 and 175 mg/day. In general, the low-potency drugs are more sedative and more hypotensive than the high-potency agents but also have less of a tendency to produce extrapyramidal side-effects. The typical antipsychotic drugs differ from one another with regard to potential for other side-effects, e.g. weight gain and hypotension, but have comparable efficacy as antipsychotic agents.[4]

Atypical antipsychotic drugs are those antipsychotic agents with a significantly lower propensity to produce EPS at clinically effective doses.[1] They are also characterized by a more diverse and complex pattern of pharmacological activity, including serotonin $(5\text{-hydroxytryptamine})_{2A}$ and dopamine D_2 antagonism as well as a variety of activities at other receptors whose contribution to their mode of action is still being elucidated.[2] Substituted benzamides, e.g. amisulpride, also have low EPS at clinically effective doses and may constitute another class of atypical agents. New classes of atypical antipsychotic drugs are emerging from research with considerable frequency at the current time.

The prototypical atypical antipsychotic drug is clozapine, a dibenzodiazepine (Table 6.2.5.1).[5] Others include aripiprazole,[6] olanzapine, quetiapine, paliperidone, risperidone, sertindole, ziprasidone and zotepine, while iloperidone,[7] asenapine,[8] and lurasidone[9] are in development and have a similar pharmacology to that of risperidone. These drugs are all more potent 5-HT_{2A} than D_2 receptor antagonists as well as multireceptor antagonists[9,10] except for aripiprazole, which is a dopamine D_2 receptor partial agonist. Bifeprunox is also a partial D_2 agonist. It lacks 5-HT_{2A} receptor blocking properties, relying instead on 5-HT_{1A} partial agonism to reduce serotonergic tone. Amisulpride and remoxipride are substituted benzamides. Both are selective D_2/D_3 antagonists.[2]

Table 6.2.5.1 Selected antipsychotic drugs and classification schemes

Drug name	Trade name	Chemical class	General class	D2 potency*
Aripiprazole	Abilify	Dihydrocarbostyril	Atypical	
Chlorpromazine	Thorazine	Phenothiazine	Typical	Low
Clozapine	Clozaril	Dibenzazepine	Atypical	
Droperidol	Inapsine	Butyrophenone	Typical	Mid
Fluphenazine	Prolixin	Phenothiazine	Typical	High
Haloperidol	Haldol	Butyrophenone	Typical	High
Loxapine	Loxitane	Dibenzazepine	Typical	Mid
Mesoridazine	Serentil	Phenothiazine	Typical	Low
Molindone	Moban	Dihydroindolone	Typical	Mid
Olanzapine	Zyprexa	Thiobenzodiazepine	Atypical	
Paliperidone	Invega	9-hydroxy metabolite of risperidone	Atypical	
Perphenazine	Trilafon	Phenothiazine	Typical	Mid
Pimozide	Orap	Butyrophenone	Typical	Mid
Promazine		Phenothiazine	Typical	Mid
Quetiapine	Seroquel	Dibenzothiazepine	Atypical	
Risperidone	Risperdal, Risperdal CONSTA	Benzisoxazole	Atypical	
Thioridazine	Mellaril	Phenothiazine	Typical	Low
Tiotixene	Navane	Thioxanthene	Typical	High
Trifluoperazine	Stelazine	Phenothiazine	Typical	Mid
Ziprasidone	Geodon	Benzisothiazole	Atypical	

* Classification on the basis of potency of D2 receptor binding for typical antipsychotic drugs only

Remoxipride was withdrawn shortly after its introduction because of a high rate of aplastic anaemia.

As will be discussed, the atypical antipsychotic drugs differ not only with regard to side-effects but also with regard to efficacy.[11,12] Atypical antipsychotic agents have been shown to have advantages, albeit modest, in treating negative mood symptoms[13–15] and to improve cognitive dysfunction in schizophrenia and perhaps other psychiatric disorders.[16–18]

Pharmacology

There is abundant evidence that dopamine plays a key role in the aetiology of psychosis and the action of antipsychotic drugs.[19] The antipsychotic action of the typical antipsychotic drugs is highly correlated with their affinities for D_2 receptors. Amphetamine and methamphetamine, which increase synaptic concentrations of dopamine, have been found to exacerbate delusions and hallucinations in some patients with schizophrenia This effect is believed to be due to stimulation of a subgroup of D_2 receptors in mesolimbic nuclei.[19,20] The cell bodies of mesolimbic dopamine neurones reside in the ventral tegmentum, the so-called A10 area, and have terminals in the nucleus accumbens, stria terminalis, and olfactory tubercle. The outflow of these regions to the thalamus and the cortex is believed to mediate psychotic symptoms. The firing rate of the mesolimbic dopaminergic neurones is subject to multiple influences, including stimulatory serotonergic input from the median raphe.[21] The origin of the dopamine neurones that terminate on cholinergic neurones in the basal ganglia is the substantia nigra, the so-called A9 region.[20] Blockade of striatal D_2 receptors in this pathway leads to the extrapyramidal side-effects produced by antipsychotic agents. A group of ventral tegmental dopamine neurones project to various regions of the cortex and comprise the mesocortical dopamine system. There is extensive evidence that these neurones are important for cognition, especially working memory,[22] as well as negative symptoms.[23] Neuroleptic drugs occupy 80 to 95 per cent of striatal D_2 receptors in patients with schizophrenia at clinically effective doses, though a lower blockade threshold of 60 per cent for improving positive symptoms has been identified.[24] Extrapyramidal side-effects occur above 80 per cent occupancy of these receptors. Blockade of D_2 receptors in the anterior pituitary gland is the basis for their ability to stimulate prolactin secretion.[25]

The prefrontal cortex has relatively low concentrations of D_2 receptors and has a higher density of D_1, D_3 and D_4 dopamine receptors.[20] The activation of D_1 receptors in prefrontal cortex may be especially critical for normal working memory and other executive type functions subserved by this brain region. However, no D1 agonists are available for treatment at the current time, although several are in development. Drugs which selectively block D_4 receptors have not been found to have an antipsychotic effect.[26] There are only limited data regarding the aetiologic or pharmacological significance of D_3 receptors in schizophrenia.

The typical antipsychotic drugs vary in their *in vitro* and *in vivo* affinities for receptors such as the dopamine D_1, histamine H_1, muscarinic, α-1 and α-2 adrenergic, and serotonergic receptors (Table 6.2.5.2), which mediate effects on arousal, extrapyramidal, cognitive, cardiovascular, gastrointestinal, and genitourinary function (Table 6.2.5.3).[27]

Thioridazine is a relatively potent antimuscarinic agent. Most of the low-potency antipsychotic agents are potent α_1 and H_1 antagonists. These affinities contribute to hypotension and weight gain, respectively. While some typical antipsychotic drugs have a high affinity for 5-HT_{2A} receptors, their affinities for D_2 receptors are even higher, which diminishes the beneficial effects of the 5-HT_{2A} receptor blockade. The specific receptor profile of each atypical antipsychotic is of special interest because it may account for critical differences among these compounds, especially in terms of side effect burden (Table 6.2.5.4).

The affinities of the atypical antipsychotic drug have been related to their efficacy and side effect profiles. As noted above, the most important determinant of atypicality for most of the currently available agents of this type is that they are more potent 5-HT_{2A} than D_2 receptor antagonists. An exception is aripiprazole, which combines potent 5-HT_{2A} antagonism and 5-HT_{1A} agonism, with partial D_2 receptor agonism. Another exception is amisulpiride, which is a selective $D_{2/3}$ antagonist with little pharmacological activity at 5-HT_{2A} receptors. Combined 5-HT_{2A} with less potent D2 antagnoism is the most consistent principle yet discovered to produce a separation between antipsychotic action and interference with motor function. This hypothesis arose from showing that it could distinguish clozapine, the prototypical atypical antipsychotic drug, and a series of other atypical antipsychotic compounds from those which have typical properties.[28] These studies suggested that the low potential for extrapyramidal side-effects of clozapine, and subsequently, olanzapine, quetiapine, risperidone, iloperidone, ziprasidone, paliperidone and asenapine are due, in part, to their relatively stronger 5-HT_2 antagonist and weak D_2 antagonist properties. The serotonin-dopamine interaction in the nigrostriatal and mesolimbico-cortical pathways appears to be mediated by stimulation of 5-HT_{2A} receptors, which are located on dopaminergic cell bodies, whereas antagonism of these receptors may release these neurones from tonic inhibition.

The atypical antipsychotic agents have the ability to increase prefrontal cortical dopaminergic activity compared with subcortical dopaminergic activity.[29] The ability to increase the release of dopamine in the prefrontal cortex may be important for atypical antipsychotic agents to improve cognition and negative symptoms. It may also contribute to decreasing the release of dopamine in the mesolimbic region, because prefrontal dopamine neurones modulate the activity of corticolimbic glutamatergic neurones that influence the release of dopamine from nerve terminals in the limbic region.[22] Typical neuroleptic drugs do not share this ability to increase dopamine efflux in prefrontal cortex. Clozapine and some of the other atypical antipsychotic drugs that are also potent 5-HT_{2A} antagonists, but not typical neuroleptics, also produce marked increases in prefrontal cortical and hippocampal acetylcholine efflux.[30] These atypical agents also produce marked increases in noradrenaline efflux in the prefrontal cortex which is correlated in time and magnitude with the increase in extracellular dopamine.[31] It is of interest that in rodents, combining ritanserin (a mixed 5-HT2a/2B/2C antagonist) or M-100907 (a selective 5-HT_{2A} antagonist) with a selective $D_{2/3}$ antagonist resulted in increased prefrontal dopamine release.[32,33] The combination of haloperidol and M-100907 also increased prefrontal dopamine release, with the greatest effects observed when lower doses of haloperidol were used.[34] Because reduced noradrenergic and dopaminergic function in prefrontal cortex and hippocampus has been

Table 6.2.5.2 Affinities of selected antipsychotic drugs at various neuroreceptors

Drug name	D2	5-HT$_{1A}$	5-HT$_{2A}$	5-HT$_{2C}$	α-1	α-2	H-1	M-1
Aripiprazole	0.95	5.6	4.6	181.0	25.0	74.0	29.0	>6K
Chlorpromazine	2.0	>3K	3.2	26.0	0.28	184.0	0.18	47.0
Clozapine	431.0	105.0	13.0	29.0	l.6	142.0	2.0	14.0
Droperidol	0.25 [173]	NA	NA	NA	NA	NA	NA	NA
Fluphenazine	0.54	145.0	7.4	418.0	6.4	314.0	7.3	>1K
Haloperidol	2.0	>1K	73.0	>10K	12.0	>1K	>3K	>10K
Loxapine	10.0	>2K	3.9	21.0	31.0	151.0	2.8	175.0
Molindone	63.0[43]	>3K[43]	320.0[43]	>10K[43]	>2K[43]	>1K[43]	>2K[43]	NA
Olanzapine	72.0	>2K	3.0	24.0	109.0	314.0	4.9	24.0
9-OH risperidone*	9.4	637.8	1.9	100.3	2.5	4.7	5.6	>10K
Perphenazine	1.4[43]	421.0[43]	5.6[43]	132.0[43]	10.0[43]	810.5[43]	8.0[43]	NA
Pimozide	0.65[43]	650.0[43]	19.0[43]	>3K[43]	197.7[43]	>1K[43]	692.0[43]	800.0[174]
Quetiapine	567.0	431.0	366.0	>1K	22.0	>3K	7.5	858.0
Risperidone	4.9	427.0	0.19	94.9	5.0	151.0	5.2	>10K
Thioridazine	10.0	108.0	11.0	69.0	1.3	134.0	14.0	33.0
Tiotixene	1.4	410.0	111.0	>1K	12.0	80.0	12.0	>10K
Trifluoperazine	1.3[43]	950.0[43]	13.0[43]	378.0[43]	24.0[43]	653.7[43]	63.0[43]	NA
Ziprasidone	4.0	76.0	2.8	68.0	18.0	160.0	130.0	>10K

All receptor binding affinities are reported as K_i (nM) using National Institutes of Mental Health (NIMH) Psychoactive Drug Screening Program (PDSP) certified data, available online at http://pdsp.cwru.edu/pdsp.php, unless otherwise specified. In general, the lower the K_i (nM) value, the higher the binding affinity for the drug at a given receptor site.

NA = human cloned receptor data not available

* 9-hydroxy (9-OH) risperidone is marketed as paliperidone

Table 6.2.5.3 Hypothesized therapeutic and adverse effects of receptor occupancy by antipsychotic drugs

Target receptor	Pharmacological activity	Therapeutic effect(s)	Adverse effect(s)
Dopamine D2	Antagonism or partial agonist effects	Reduction of positive symptoms	Extrapyramidal effects (EPS) Hyperprolactinemia
Serotonin (5-HT) $_{1A}$	Full or partial agonist effects	Cognitive enhancement Reduction of mood and anxiety symptoms	
5-HT $_{2A}$	Antagonism	Reduction of negative symptoms Reduction of EPS Reduction of mood and anxiety symptoms Increased deep sleep	
5-HT $_{2C}$	Antagonism	Reduced anxiety symptoms	Weight gain
Adrenergic α-1	Antagonism		Orthostatic hypotension Dizziness
Adrenergic α-2	Antagonism		Reflex tachycardia
Histamine H-1	Sedation	Sedation Drowsiness Weight gain	
Muscarinic (cholinergic) M-1	Antagonism	Reduction of EPS	Blurry vision Exacerbation of acute angle closure glaucoma Sinus tachycardia Constipation Urinary retention Memory dysfunction

Adapted from Kelly, D.L. and Love, R.C. Ziprasidone and the QTC interval: pharmacokinetic and pharmacodynamic considerations, Psychopharmacology Bulletin, **35**, 66–79, copyright 2001, MedWorks Media Global, LLC.

Table 6.2.5.4 Adverse effects of selected antipsychotic drugs

Typical antipsychotic drugs

	EPS	Tardive dyskinesia	Prolactin elevation	Sedation	Weight gain	Orthostasis	Anti-cholinergic	Diabetes exacerbation & dyslipidemia
Chlorpromazine Fluphenazine Haloperidol Loxapine Mesoridazine Molindone Perphenazine Thioridazine Tiotixene Trifluoperazine	Some (for low potency* drugs) - +++ (for high-potency* drugs)	++ - +++	++ - +++ (risk higher for high-potency drugs)	Some (for high potency drugs) - +++ (for low-potency drugs); ? least for molindone	Some (for high potency drugs) - +++ (for low-potency drugs); ? least for molindone	Some (for high potency drugs) - +++ (for low-potency drugs)	Some (for high potency drugs) - +++ (for low-potency drugs)	+ - ++

* See Table 6.2.5.1 for list of low-, mid-, and high-potency (with respect to dopamine D2 receptor blockade) antipsychotic drugs

Adapted from the International Psychopharmacology Algorithm Project (IPAP) algorithm for the treatment of schizophrenia, available at www.ipap.org, copyright 2008 International Psychopharmacology Algorithm Project (IPAP)

Atypical antipsychotic drugs

	EPS	Tardive dyskinesia	Prolactin elevation	Sedation	Weight gain	Orthostasis	Anti-cholinergic	Glucose dysregulation & dyslipidemia
Amisulpride	+	Rare	+++	+	0 - +	+	0	0
Aripiprazole	0 - +	0 - +	0	0 - +	0 - +	+ - ++	0	0
Clozapine	0	0	Transient	+++	+++	+++	+++	+++
Olanzapine	0 - + (if < 10 mg/day)	Rare	+ (if < 20 mg/day)	++	+++	+	+	+++
Quetiapine	0	Rare	0	++	+ - ++	++	0 - +	++
Risperidone	+ (less if < 4 mg/day)	Rare	+++	+	+ - ++	++	0	+
Ziprasidone	0 - +	Rare	0 - +	0 - ++	0	+ - ++	0	0

Sufficient data for paliperidone, iloperidone and asenapine are not yet available for inclusion in this table.

Adapted from the International Psychopharmacology Algorithm Project (IPAP) algorithm for the treatment of schizophrenia, available at www.ipap.org, copyright 2008 International Psychopharmacology Algorithm Project (IPAP)

associated with negative symptoms and cognitive impairment in schizophrenia,[22,35] the cortical release of these two neurotransmitters, and possibly also acetylcholine, may provide a pharmacological basis for the advantages of atypical antipsychotics over typical neuroleptic drugs in the treatment of these critical symptom domains. In patients with schizophrenia who were stabilized on typical neuroleptics, the addition of mianserin, a 5-$HT_{2A/C}$ and adrenergic α-2 antagonist, was associated with improved neurocognitive performance,[36] adding further support to a role of 5-HT_{2A} receptors in the treatment of cognitive dysfunction in schizophrenia.

The importance of serotonin receptors other than 5-HT_{2A} for the action of antipsychotic drugs has received considerable attention. Activation of 5-HT_{1A} receptors are believed to have a dopamine modulating effect similar to that of 5-HT_{2A} antagonism.[37] Under experimental conditions, 5-HT_{1A} agonists have been shown to stimulate cortical dopamine release[38,39] and, in schizophrenic patients who were stabilized on haloperidol, the addition of tandospirone, a 5-HT_{1A} partial agonist, resulted in improved neurocognitive performance.[40] This effect has also been demonstrated more recently for buspirone, another 5-HT_{1A} partial agonist.[41] Serotonin-1A receptors may be important for cognitive effects of at least some of the atypical antipsychotic drugs that are active at this receptor site. Activity at 5-HT_{1A} receptors is not shared by all antipsychotic drugs (Table 6.2.5.2), however. Antagonism of 5-HT_{2C} receptors also appears to result in cortical dopamine and norepinephrine release, as well as in the nucleus accumbens.[42] The cognitive effects of selective 5-HT_{2C} antagonists added to typical neuroleptic drugs in patients with schizophrenia have not been examined. As is the case with 5-HT_{1A} activity, not all atypical antipsychotic drugs are active at 5-HT_{2C} receptors (Table 6.2.5.2). Like antagonism at histamine H_1 receptors,[43] 5-HT_{2C} antagonist activity may be related to antipsychotic induced weight gain.[44]

Atypical antipsychotics may display regional selectivity in terms of their dopaminergic activity, relative to typical neuroleptics. For instance, atypical antipsychotic drugs appear to preferentially block cortical D_2 receptors, relative to those located in the striatum.[45,46] Haloperidol results in proportionally equivalent D_2 blockade in both brain regions.[47] The atypical antipsychotics also increase the expression of the early intermediate gene c-*fos*, in the prefrontal cortex and the shell of the nucleus accumbens, while sparing the core of the latter region and the striatum. Typical neuroleptic drugs have the opposite effect on c-*fos* expression. Sparing the dorsal

striatum is believed to be related to the low potential for extrapyramidal side-effects of these agents.[2,21]

Clozapine, olanzapine, risperidone, and quetiapine are able to block the interference in prepulse inhibition produced by d-amphetamine, apomorphine, or phencyclidine at doses that do not interfere with locomotor function. Clozapine and M100907 are able to block the effects of phencyclidine, an *N*-methyl-D-aspartate receptor antagonist, on locomotor activity in rodents. This suggests the ability of rat 5-HT_{2A}-receptor blockade to block some of the effects of phencyclidine which is one of the more important models for schizophrenia.[2,21] In a recent single photon emission tomography (SPECT) study, patients with schizophrenia who received treatment with clozapine evidenced reduced NMDA-active radiotracer binding compared with healthy controls, drug free patients with schizophrenia, and patients with schizophrenia who were treated with typical neuroleptics.[48] The extent of involvement of other atypical antipsychotic drugs relative to typical antipsychotics at NMDA receptors and other glutamatergic targets is an area of active interest. Other receptor targets that are of special interest in terms of improving cognitive functioning and selected psychotic symptoms include M1 muscarinic, α-7 nicotinic, and α-1 and α-2 adrenergic receptors.

Administration, pharmacokinetics, and dosage

Administration

(a) Typical antipsychotic drugs

The major uses of the antipsychotic drugs are for the treatment of schizophrenia, mood disorders typically with psychotic features, and senile psychoses.[4,49] Other indications are discussed elsewhere in this book in the consideration of the management of specific disorders, such as Tourette's syndrome, and aggression. The major advantage of the typical neuroleptic drugs is their ability to improve positive symptoms, i.e. delusions and hallucinations. Administration of typical neuroleptic drugs leads to the complete or nearly complete elimination of positive symptoms and disorganization of thought and affect in about 60 to 70 per cent of patients with schizophrenia and an even higher proportion of those with psychotic mania and psychotic depression.[49] The antipsychotic response in schizophrenia and mania is sometimes apparent within a few days in many patients but usually takes up to several weeks or months. A reasonable duration for a therapeutic trial with one of these agents is 4 to 6 weeks. It is not appropriate to switch medications after 1 or 2 weeks, even if a response is not apparent, unless side-effects pose a serious problem. Positive symptoms (delusions and hallucinations) do not respond to typical neuroleptic drugs in about 10 per cent of schizophrenic patients even during the first episode.[50] Another 20 per cent of patients with schizophrenia develop resistance to these agents during the subsequent course of their illnesses.[51] Development of resistance to typical neuroleptic drugs may occur at any time during the course of treatment, even after many years of control of positive symptoms. Such patients are more likely to respond to clozapine[51] or one of the other atypical antipsychotics.[50,51]

The average doses of the typical neuroleptic drugs are given in Table 6.2.5.5. The best results with these drugs in terms of efficacy and side-effects may be expected with the lowest dose needed to produce control of positive symptoms with the fewest extrapyramidal side-effects.[4,49]

There are some patients for whom higher doses are indicated, but most controlled studies have failed to find benefits from high-dose strategies of combining two or more of these agents. Increasing the dose of these agents when patients fail to respond rapidly, for example within days, is not recommended. Augmentation with a benzodiazepine may be useful to decrease anxiety until the lower doses of neuroleptic drugs produce adequate control of positive symptoms.[4,49] Patients who may require higher doses of neuroleptic drugs to respond adequately are at greater risk of hyperprolactinaemic effects, EPS, and tardive dyskinesia and are generally better treated with an atypical antipsychotic drug.

However, the improvement in positive symptoms which is often achievable with the typical antipsychotic drug is only one element in the treatment of schizophrenia and is not sufficient grounds for judging response to be adequate. Additional efficacy factors of major importance are summarized in Table 6.2.5.6.

Tolerability and safety factors, such as compliance, tardive dyskinesia, weight gain, and medical morbidity are also major elements in outcome and are influenced by the choice of a typical or atypical antipsychotic drug. Typical neuroleptic drugs are not as effective for improving primary negative symptoms of schizophrenia in the majority of patients.[52,53] There is a consensus that typical neuroleptic drugs can improve negative symptoms that are secondary to positive symptoms and depression while at the same time possibly causing secondary negative symptoms due to their ability to produce extrapyramidal side-effects.[52] Abnormalities in specific domains of cognition (Table 6.2.5.6) are present in first-episode schizophrenic patients at a moderate to severe level and show slight to moderate, rarely severe, deterioration during the course of illness.[54,55] Approximately 85 per cent of patients with schizophrenia are clinically impaired in one or more domains of cognition.[55,56] Cognition has been shown to be perhaps the most critical determinant of functional capacity among patients with schizophrenia, even more so than positive symptoms.[57] Typical neuroleptic drugs usually do not improve cognitive function.[58] Those typical neuroleptic drugs such as thioridazine and mesoridazine, which have strong antimuscarinic properties, may produce further impairment in some memory functions.[58]

All of the typical neuroleptic drugs are likely to be equally effective in treating either the initial presentation or recurrent psychosis due to breakthrough of symptoms, despite compliance, or because of having stopped medication[4,49,51] First-episode patients with schizophrenia usually require much lower doses than patients with two or more episodes, suggesting some progression of the disease process or development of tolerance to the mechanism of action of these drugs.[59] Doses for more chronic patients should be in the range of 5 to 10 mg haloperidol equivalents per day (Table 6.2.5.5) for up to 4 to 6 weeks unless there is a major need for chemical means to prevent harm to self or others, to decrease excitement, or induce sleep.[60] Auxiliary medications for anxiety and sleeplessness, for example benzodiazepines, may supplement these low doses of antipsychotics.[61]

Parenteral injections of haloperidol, chlorpromazine, or other neuroleptics may be needed for patients who refuse oral medication or where very rapid onset of action is needed to control acutely dangerous behaviours if less restrictive means either fail or cannot be utilized safely. Commonly, haloperidol (2–10 mg) with or without lorazepam (2–4 mg) is delivered intramuscularly every 30

Table 6.2.5.5 Oral dosing of antipsychotic drugs

Typical antipsychotic drugs				
	Equivalent doses (mg/day)	**Starting dose**	**Titration schedule**	**Dose range (mg/day)**
Chlorpromazine[a]	100	15–50 mg BID-QID	As clinically indicated	300–1000 (divided QD-QID)
Fluphenazine[b]	2	0.5–10 mg/day (divided Q6–8 hours)	As clinically indicated	5–20
Haloperidol[c]	2	0.5–5 mg BID	As clinically indicated	5–20
Loxapine	10	10 mg BID	As clinically indicated	30–100
Mesoridazine	50			150–400
Molindone	10	50–75 mg/day divided TID-QID	As clinically indicated	30–100
Perphenazine[d]	10	4–8 mg TID (8–16 mg BID-QID if hospitalized)	As clinically indicated	16–64
Thioridazine	100	50–100 mg TID	As clinically indicated	300–800
Tiotixene	5	2 mg TID	As clinically indicated	15–50
Trifluoperazine	5	2–5 mg BID	As clinically indicated	15–50

For elderly patients, or those with renal or hepatic problems, doses of drug may need to be reduced by one-half or more

[a] Short-acting IM formulation may be given 25–50 mg (may repeat after 1–4 hrs as required); may gradually increase dose up to 400 mg IM Q 4-6 hrs (maximum of 2000 mg/day) may be needed for severe cases

[b] Short-acting IM formulation may be given 2.5–10 mg/day in Q6–8 hr intervals; Depot IM formulation may be given 12.5–25 mg Q 3 weeks

[c] Short-acting IM formulation may be given 2–5 mg Q 1–4 hrs; Depot IM formulation may be given at approximately 10–20 times the stable oral dose Q 4 weeks

[d] Short-acting IM formulation may be given 5–10 mg Q 6 hrs (maximum of 30 mg/day)

Atypical antipsychotic drugs			
	Starting dose	**Titration schedule**	**Dose range (mg/day)**
Aripiprazole[a]	10–15 mg daily	As clinically indicated, every 2 weeks	10–30
Clozapine	12.5 mg QD-BID	Increase by 25–50 mg/day until usual effective dose of 300–450 mg/day after 2–4 weeks	150–600
Olanzapine[b]	5–10 mg daily	As clinically indicated, by 5 mg/day every 7 days	10–30
Paliperidone	6 mg/day	As clinically indicated, by 3 mg/day Q 2–4 week increments, up to 12 mg daily	6–12
Quetiapine	25 mg BID	Increase by 25–50 mg BID-TID on days 2 and 3, to target dose of 300–400 mg daily (QD – TID) by day 4. Further increases as clinically indicated by 25–50 mg BID every 2 days.	300– 800
Risperidone[c]	0.5–1 mg BID	Increase by 0.5–1 mg BID on days 2 and 3, with further dose increases thereafter by 0.5–1 mg increments Q 7 days as required	2–8
Ziprasidone[d]	20 mg BID with food	Increase by 20–40 mg BID every 2 days to target dose of 80 mg (all doses with food)	120–200

For elderly patients, or those with renal or hepatic problems, doses of drug may need to be reduced by one-half or more, and titration may be slower

[a] Short-acting IM formulation may be given at 9.75 mg, though the lower 5.25 mg dose may be indicated in some situations.

[b] Short-acting IM formulation may be given 10 mg as required (may be repeated after 2 hrs, up to 30 mg/day).

[c] Long-acting IM formulation may be initiated at 25 mg Q 2 weeks (continue oral risperidone dose for 3 weeks), with increases as clinically indicated every 4 weeks up to a dose of 50 mg Q 2 weeks

[d] Short-acting IM formulation may be given 10–20 mg as required (may be repeated Q 2–4 hrs as needed, up to 40 mg/day)

to 60 minutes as required, up to three doses. Doses of haloperidol given intramuscularly in such situations generally should not exceed 18 mg per day. Oral medication should be substituted as soon as feasible. If positive symptoms fail to respond to a single trial of a typical neuroleptic drug at adequate doses in patients with schizophrenia, there is evidence that switching to another typical antipsychotic, even of a different chemical class, is unlikely to produce greater control.[4,49,51] This is likely to be true for other indications for the use of antipsychotic agents as well.

In cases of repeated illness relapse due to poor compliance or when patients prefer it, the use of long acting (e.g. depot) injectable antipsychotic medications, typically administered once every 2–4 weeks, may be used. The use of injectable antipsychotic medication has been associated with lower rates of relapse and rehospitalization and greater global improvement compared with oral typical neuroleptics,[62] possibly as a result of ensured drug delivery. Long acting injectable drugs should not be given to ameliorate acute behavioural disturbances.

(b) Atypical antipsychotic drugs

As implied above, there are major advantages for many patients to be treated with the atypical antipsychotic drugs and it is generally

Table 6.2.5.6 Target signs and symptoms for the pharmacological management of schizophrenia

Target	Description	
Positive syndrome	Hallucinations Delusions	◆ Typically the most amenable to treatment with all antipsychotic drugs
Negative syndrome	Avolition Apathy Anhedonia Lack of responsiveness Poor rapport with others Passive social withdrawal Poverty of speech Affective flattening	◆ Robustly correlated with functional impairment in schizohprenia ◆ More difficult to treat pharmacologically, and may required longer to respond than positive signs and symptoms ◆ Pharmacological adjuncts may be needed, though under-studied ◆ Atypical antipsychotic drugs are believed to be more efficacious than typical neuroleptics
Hostility/excitement	Verbal or physical aggression	◆ Typically amenable to treatment with all antipsychotic drugs ◆ Use of parenteral formulation may be required
Mood and anxiety symptoms	Depressed mood Anxious mood Nervousness Panic symptoms Suicidal ideation	◆ Believed to be more responsive to treatment with atypical antipsychotic drugs ◆ Clozapine has demonstrated superiority for treating chronic suicidality in schizophrenia
Cognitive impairment (psychopathological definition)	Disorientation Problems with abstraction Attentional problems Preoccupations Disorganized thought processes	◆ Some domains respond favourably to antipsychotic drug treatment, though response is often incomplete
Cognitive impairment (neuropsychological testing definition)	Working memory Attention/vigilance Verbal learning/memory Visual learning/memory Problem solving Processing speed	◆ Neuropsychological deficits, like negative signs and symptoms, are robustly correlated with functional outcome in schizophrenia ◆ Very difficult to treat with medication alone ◆ Atypical antipsychotic drugs are believed to be superior to typical neuroleptics, though effect sizes are only mild to moderate for the former

recommended that, where possible, these agents be considered as the first-line treatment.[63,64] The atypical antipsychotic drugs are the dominant antipsychotic treatment for schizophrenia, mania, and psychotic depression in clinical practice in many parts of the world. However, there is considerable international variation in their usage. Cost factors may explain part of the variance in the use of these agents within and between countries. The typical neuroleptic drugs are no longer covered by patent protection and are available in inexpensive generic forms. There are a number of patients whose psychosis is adequately controlled by these agents and they (and their families and prescribers) are content to continue them even when informed of the potential advantages of the newer antipsychotic agents. When only the cost of medication is considered, it may seem that fiscal reasons argue for continuation of typical neuroleptic drug treatment since the atypical agents can cost up to 100 times more. In addition, the widely accepted notion of greater overall benefit from treatment with atypical antipsychotic drugs, as opposed to typical neuroleptics, was recently challenged by results from two effectiveness studies. The first of these demonstrated no significant difference in all-cause discontinuation from the study as the primary endpoint, as well as discontinuation for lack of efficacy, between atypical drugs, with the exception of olanzapine, and the typical neuroleptic perphenazine.[65] The latter study reported a lack of significant differences in quality of life between patients who received naturalistic treatment with typical or atypical antipsychotic drugs.[66] Methodological limitations, detailed discussion of which is beyond the scope of this chapter, limit the conclusions that can be drawn from these reports about the relative merits of one class of antipsychotics versus another, both of which are in disagreement with the majority of the clinical literature that documents differences between these broad classes of antipsychotic drugs across a wide range of outcomes. Because medication costs account for, at most, 5 per cent of the total costs of schizophrenia, with the major costs being hospitalization and indirect costs such as lost income and disability income to support patients in the community, more effective and tolerable medications may offset their greater cost.[67,68] As such, atypical antipsychotic drugs are recommended as first line treatments of schizophrenia and related psychotic disorders. Each will be discussed separately.

(i) Clozapine

Clozapine was synthesized in 1959 as part of a project to discover antipsychotic drugs with low potential for extrapyramidal side-effects. It proved to be one of the most interesting and clinically important compounds ever discovered. It was labelled as atypical because of its ability to block amphetamine-induced locomotor activity, one of the most widely accepted models for antipsychotic activity, without producing catalepsy in rodents, the leading model for causation of extrapyramidal side-effects in humans. Subsequent clinical studies showed it to have the lowest extrapyramidal side-effects of any antipsychotic drug known.[5,69] Clinical trials in the 1960s and 1970s suggested it was also superior in

efficacy with regard to control of positive symptoms, but given the standards of clinical trials of that era, these conclusions could not be relied upon.[70] In 1975, 6 years after its introduction in Europe, clozapine's ability to cause granulocytopenia or agranulocytosis was first reported. Six deaths occurred in clozapine-treated patients in a geographically restricted area of Finland over a short period of time. The role of clozapine in these deaths is still uncertain because no other such clustering has ever occurred in Finland, or elsewhere. Nevertheless, clozapine was withdrawn from general use, although it remained available for humanitarian use in patients who had previously received it, for individual cases where it seemed indicated because of its low potential for extrapyramidal side-effects, and for research purposes.[69]

Clozapine was reintroduced in 1989 after it was demonstrated to be superior to chlorpromazine to improve positive and negative symptoms in 300 patients who were resistant to the action of at least three typical neuroleptics.[71] Thirty per cent of the patients treated with clozapine responded after 6 weeks of treatment compared to 4 per cent of the chlorpromazine-treated patients. Subsequent studies have shown that up to 60 to 70 per cent of patients will respond within 6 months of treatment. Patients with shorter duration of illness tend to respond better. Some predictors of response include weight gain and absence of atrophy in the prefrontal cortex.[52] Clozapine has been reported in several studies to reduce the risk of suicide.[52,72] It has been shown in a large number of studies to improve some aspects of cognitive function, especially verbal fluency, immediate and delayed verbal learning and memory, and attention.[16–18]

Because of the side-effect profile of clozapine, it is not generally used as a first-line drug. On the other hand, monitoring the white blood count for the development of agranulocytosis or granulocytopenia, as well as improved methods of treating agranulocytosis, have made it much safer to use. Clozapine is still probably underutilized in many parts of the world. Any patient with an unsatisfactory response to the typical neuroleptics and at least one atypical antipsychotic should be considered for clozapine treatment. This amounts to at least 20 per cent of schizophrenics. Clozapine use may also be considered for patients with schizophrenia who are at high risk for suicide, even if the aforementioned threshold of inadequate response to other drugs has not yet been met.

Clozapine is usually given twice daily, but sometimes more than half of the dose or the entire dose is given at sleep time to minimize daytime sedation. The daily dosage is gradually titrated to the target range described in Table 6.2.5.5. Patients who are treatment resistant may require higher doses. Typical or non-clozapine atypical antipsychotic drugs should be discontinued either before beginning clozapine or by eliminating them over a 1- to 2-week period as the dose of clozapine is increased. Because clozapine produces only about 40 to 50 per cent occupancy of striatal D_2-receptors,[73] and some of its key advantages are believed to be related to its low D_2-receptor blockade, concomitant administration of typical neuroleptic drugs would be predicted to interfere with some of the benefits of clozapine and, thus, should not ordinarily be prescribed with clozapine. However, some patients with persistent positive symptoms despite an adequate trial of clozapine monotherapy might be expected to benefit from the addition of low-dose haloperidol, or its equivalent, to provide additional low level D_2-receptor blockade.

Determination of clozapine plasma levels is useful whenever patients are not responding adequately. If response is inadequate, various approaches to augment response have also been utilized. In addition to adding a low dose of a typical neuroleptic, as mentioned above, it may be useful to augment clozapine treatment with valproic acid or other mood stabilizer (such as lithium, carbamazepine, lamotrigine or topiramate), anxiolytic drugs, or an antidepressant.[52] The choice of augmenting agent is largely driven by symptomatic considerations, or pharmacokinetic interactions in the case of fluvoxamine. However, none of these strategies have strong empiric support. One exception may be the addition of sulpiride, which may result in a significant reduction in symptom burden when added to clozapine.[74] Electroconvulsive therapy (ECT) also resulted in a modest further reduction in symptoms when used in conjunction with clozapine, and appears to be well tolerated.[75] It is difficult to postulate a rationale for adding another atypical antipsychotic, with the exception of amisulpride, because of their similarity in pharmacology to clozapine. It should be discontinued if side-effects are intolerable, or if there is no apparent response after a 6-month trial of clozapine alone and subsequent trials with augmentation therapy. Clearly, further studies involving clozapine partial- or non-responders are urgently needed. It should be noted that discontinuation of clozapine can precipitate a severe relapse even when clozapine is slowly tapered.[76]

(ii) Risperidone

Risperidone is useful as a first-line drug for the treatment of all forms of schizophrenia, including residual schizophrenia.[77–79] Definitive data are lacking for its efficacy in patients who are neuroleptic resistant or who have failed to respond to other atypical antipsychotics, including clozapine.[80] Clinical experience is not supportive of widespread efficacy in these groups but there may be some responders. However, risperidone may be useful in patients who fail to tolerate other antipsychotic agents because of side-effects not shared by risperidone, such as anticholinergic effects. Risperidone is well-tolerated in low doses by the elderly and has been widely used in the United States for the treatment of a variety of senile psychoses.[81,82] Its efficacy against haloperidol was established in a series of multicentre trials which demonstrated advantages for risperidone in overall psychopathology in mainly chronic schizophrenic patients in an acute exacerbation at doses in the 6 to 8 mg/day range.[11,77] However, these doses have proven to be higher than is needed for most patients in clinical practice, possibly reflecting some of the problems in generalizing from controlled clinical trials. The doses for schizophrenia most often used in non-elderly adults are now 4 to 6 mg/day. First-episode patients may not tolerate higher doses (e.g. above 5 mg per day), and some may respond to as little as 1 to 2 mg/day. Some treatment-resistant patients may need doses higher than 6 mg/day; however, the results of clinical studies in this population have been mixed. Whether prolonged trials, i.e. up to 6 months, are useful in such patients, as they are in clozapine patients, is not yet known.

Beyond treatment of acute symptoms, the position occupied by risperidone as a first line treatment option is also supported by long term maintenance phase and relapse prevention studies. For instance, relative to haloperidol, risperidone has also been associated with a lower risk of relapse (34 vs. 60 per cent) over a

minimum of 12 months of treatment.[83] In another study that |retrospectively compared rates of rehospitalization for patients who received treatment with risperidone, olanzapine, or typical neuroleptics, rehosptialization rates for risperidone and olanzapine were similar, and both were significantly less than those of patients treated with typical neuroleptics.[84]

Risperidone is usually initiated at low doses (e.g. 1–2 mg daily) and is titrated into the dosage range provided in Table 6.2.5.5. The medication is often initiated in twice daily dosing; however, because its primary active metabolite, 9-OH risperidone, is pharmacologically equivalent to its parent drug and because it has a longer elimination half-life, once daily dosing is also possible. Risperidone is available in soluble wafer and liquid forms, which may be advantageous for patients who have swallowing difficulties or require taking their medication in a non-pill form for other reasons, including their own preference.

For patients who have a history of poor compliance leading to frequently relapsing illness, or for those who prefer it, a long acting injectable form of risperidone is available (Table 6.2.5.1) for administration typically every two weeks. For treatment responsive patients, response may be expected to occur in the 25–50 mg (per every two week dose) range,[85] however, oral risperidone must be continued through at least the first 3 weeks of treatment with the long acting injectable form before being slowly tapered. Supplementation with oral medication may be required when the dose of the long-acting drug is upwardly adjusted due to breakthrough psychotic symptoms. As is the case with long-acting injectable typical antipsychotics, long-acting injectable risperidone should not be used acutely to control dangerous behaviours.

Risperidone has more of a tendency to produce extrapyramidal side-effects than any of the other atypical antipsychotics but this can be minimized by using the lowest dose which controls positive symptoms and adding an anticholinergic drug, if necessary.[82] Addition of a typical neuroleptic to risperidone will increase the risk of extrapyramidal side-effects. Risperidone is not well tolerated by patients with Parkinson's disease because of extrapyramidal side-effects. There are some data suggesting the risk of tardive dyskinesia in patients with schizophrenia, and especially the elderly, with risperidone is less than that of the typical neuroleptic drugs.[86]

Among atypical antipsychotic drugs, risperidone and paliperidone appear to be the most liable in terms of increasing prolactin release. As is the case with EPS, the effect of risperidone on prolactin concentration appears to positively correlate with dose.[25] The changes may occur in both men and women; however, the greatest elevations appear to occur among women. Elevations in prolactin levels as a result of treatment with risperidone do not always translate into clinical symptoms such as sexual dysfunction or gynaecomastia in men and menstrual changes and breast discharge in women; however, patients should be monitored clinically for these effects, and prolactin concentrations measured if these symptoms occur.

The issue of whether the improvement in negative symptoms by risperidone and other atypical antipsychotic drugs is due to an effect on so-called primary negative symptoms versus secondary negative symptoms has been much debated. Data from large multicentre trials of risperidone versus clozapine show an effect on primary negative symptoms as residual change left after adjusting for improvement due to decreases in positive or depressive symptoms and extrapyramidal side-effects.[87] In addition, results from a meta-analysis of 6 studies comparing risperidone to typical neuroleptic treatment indicated greater response rate for negative symptoms (defined as achieving >20 per cent reduction in negative symptom burden) as well as greater reduction in anxious/depressive symptoms among risperidone treated patients.[88] Risperidone has a greater ability to improve cognition in schizophrenia than the typical neuroleptic drugs.[17] Improvement in working memory has been the strongest finding, while improvements in attention, executive function, and verbal learning and memory have also been reported. Risperidone has been shown to be a cost-effective treatment for schizophrenia,[89] especially in its long-acting injectable form,[90] and to improve quality of life,[91] firm conclusions about relative cost-effectiveness between atypical antipsychotic drugs for non-treatment-resistant schizophrenia are difficult to draw at this time. Further research in this important area is needed.

In summary, risperidone is a first line pharmacological treatment of schizophrenia and other forms of psychosis. It may produce significant advantages over typical neuroleptic drugs with regard to negative symptoms, cognition, and extrapyramidal side-effects, but it does produce dose dependent increases in EPS risk, and increases in serum prolactin levels resembling those of typical antipsychotic drugs. It should be used at the lower doses where possible. A long-acting injectable form of this medication should be considered a first line treatment option in cases of frequent relapse due to poor medication compliance.

(iii) Olanzapine

Olanzapine is indicated as a first-line treatment for all forms of schizophrenia[63,92] with the caveat that it has not been shown to be as effective as clozapine in neuroleptic-resistant patients at conventional doses.[8,60] However, some patients of this type do respond to olanzapine,[93] perhaps at high doses.[94,95] There are no means yet to determine which of this group of patients will respond to olanzapine (or risperidone) so some clinicians may elect a trial with either of these agents before considering clozapine.

The efficacy of olanzapine in treating psychosis and negative symptoms in patients with an acute exacerbation of schizophrenia has been firmly established in a variety of large-scale, multicentre trials.[92] In these trials, olanzapine at doses of 10 to 20 mg/day has been superior to placebo and equivalent or superior to haloperidol in some measures of total psychopathology, positive, or negative symptoms. For example, in the North American multicentre trial, high-dose olanzapine (15 ± 5 mg/day) was superior to haloperidol (15 ± 5 mg/day) in the treatment of negative symptoms.[96] The effect of olanzapine to improve negative symptoms was found to be on primary rather than secondary negative symptoms.[97]

Olanzapine has also been found to be effective as a maintenance treatment of schizophrenia.[98] The estimated relapse rates, defined as the need for hospitalization, during a 1-year period in three studies of patients receiving olanzapine for maintenance treatment were 19.6 to 28.6 per cent. These rates were significantly lower than those in patients receiving placebo, ineffective doses of olanzapine, or haloperidol.[98] Olanzapine has some efficacy in treating anxious and depressive symptoms,[99] as well as cognitive dysfunction,[17,18] associated with schizophrenia or schizoaffective disorder. Pharmacoeconomic studies and investigations of medication effects on quality of life measures indicate that olanzapine has

a beneficial cost-outcome profile. For instance, in one investigation, the higher cost of olanzapine relative to haloperidol was offset by olanzapine-treatment associated reductions in rehospitalization and overall treatment costs.[100] Olanzapine treatment has also been associated with better outcomes as assessed by overall and health-related quality of life relative to haloperidol.[101] As mentioned above, olanzapine was found to be the most effective antipsychotic drug in the recent Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) phase I study.[65]

The average clinical dose of olanzapine is 12.5 to 20 mg/day but many patients may be expected to respond to lower doses (e.g. 10 mg daily).[102] A principle advantage of olanzapine is its once daily dosing and the feasibility of starting the medication at a dose that is clinically effective for most patients. Doses higher than 20 mg/day are rarely more effective than lower doses, especially for non-refractory cases. Augmentation of olanzapine with typical neuroleptic drugs or risperidone should be done sparingly to avoid extrapyramidal side-effects and possibly compromising efficacy. Olanzapine is also available in a soluble wafer form that may be preferred to the pill form by some patients, especially those with swallowing difficulties and related problems.

For acute situations where rapid control of agitation, hostility or other dangerous behaviours is required, olanzapine is available as a short-acting injectable medication (Table 6.2.5.1).[103] Similar to short-acting typical neuroleptic drugs, the medication is delivered intramuscularly. Doses of 5 to 10 mg per injection may be given, depending on the severity of the target behaviours. A long-acting formulation is in clinical testing.

In summary, olanzapine has found wide acceptance as an atypical antipsychotic drug because of its once-a-day administration, efficacy for negative symptoms, improvement in cognitive function, and low extrapyramidal side-effect profile. Significant weight gain and other metabolic effects may be a problem for some patients, as will be discussed below.

(iv) Quetiapine

Quetiapine has been shown to be as effective as typical antipsychotics, with fewer extrapyramidal side-effects and no effect on serum prolactin levels.[104,105] Part of the reason for this may be that quetiapine and clozapine both appear to bind more loosely to striatal D2 receptors than other antipsychotic drugs, and that both drugs show antipsychotic activity at D2 receptor occupancies that are well below the 60 per cent threshold identified for most other antipsychotic drugs.[106] In spite of this similarity with clozapine, quetiapine does not appear to have efficacy comparable to clozapine for treatment-resistant patients.

The efficacy of quetiapine for acute phase schizophrenia is supported by results from several randomized, controlled trials that documented superiority of quetiapine relative to placebo across several doses, with some patients with some patients responding to 150 mg/day and others requiring 750 mg/day.[104] For instance, in one high- (750 mg/day) versus low-dosage (250 mg/day) study, both dosage groups evidenced greater reduction in positive symptoms relative to placebo; however, the differences were significant only for the high-dose group.[105] In another study that assessed multiple fixed doses of quetiapine (75 to 750 mg/day) compared with haloperidol and placebo, significant differences in improvement over placebo for quetiapine were observed in the dosage range of 150 to 750 mg/day.[104]

Quetiapine's effect on negative symptoms continue to be investigated. One placebo controlled comparison documented improvements in negative symptoms with quetiapine treatment across a wide range of doses, with the greatest improvement reported at 300 mg daily.[104] In the high- vs. low- dose study reviewed above, the high-dose group also experienced greater improvement in negative symptoms relative to placebo.[105] Like risperidone and olanzapine, quetiapine appears to improve depressive symptoms[107] and certain cognitive deficits[17,18] associated with schizophrenia or schizoaffective disorder. The improvements in cognition with quetiapine appear to be superior to those of haloperidol.[108]

These results suggest that, overall, the greatest improvement in positive and negative symptoms may occur when quetiapine is used at the higher end of its dosage range. The average clinical dose appears to be between 300 and 500 mg/day, usually given twice daily though some benefit from the medication when given only once daily. The effects of using higher doses for patients who do not respond adequately to these doses are uncertain. A titration of the dosage is required after initiating the medication. From the viewpoint of EPS and hyperprolactinaemic effects, quetiapine appears to confer only low risk. As such, it, like clozapine, appears to be well tolerated even among patients with idiopathic Parkinson's disease.[109] Sedation may be a limiting side effect for some, especially during dosage titration. Weight related, metabolic, and other adverse effects will be discussed in greater detail below.

In summary, quetiapine also appears to be effective for a wide range of schizophrenia-associated symptoms and confers a lower level of risk in terms of antidopaminergic adverse effects. The dosage range of this medication may be quite wide, though patients may have a greater chance of benefiting from the medication at the higher end of this range.

(v) Ziprasidone

Ziprasidone has a varied receptor occupancy profile. Like most atypical antipsychotic drugs, it displays high affinity 5-HT2A binding coupled with relatively lower affinity D2 receptor binding. Ziprasidone is also a 5-HT1A agonist, as well as both a serotonin and norepinephrine reuptake pump inhibitor.[110] This profile predicts a wide range of pharmacological activity against core psychotic symptoms, negative and affective symptoms, as well as neurocognition.

Ziprasidone, like quetiapine, has been shown to be superior to placebo for the reduction of total psychopathology and positive and negative symptoms.[111,112] There is limited evidence to suggest superiority over typical neuroleptics with regard to improvement in positive and negative symptoms.[111,113] Studies of multiple fixed doses of ziprasidone vs. haloperidol at conventional doses indicate that ziprasidone yields similar efficacy to haloperidol for reducing positive symptoms and global psychopathology at a dose of 160 mg/day.[113] Doses greater than 160 mg/day have not been systematically investigated.

Ziprasidone significantly improved negative symptoms and reduced the risk of relapse compared to placebo in a 1-year maintenance study in stable hospitalized chronic schizophrenic patients.[114] These maintenance phase effects were not dependent on the daily dose of ziprasidone. In a 28-week comparison with haloperidol, the two groups evidenced similar overall effects for positive symptoms; however, between groups differences were

documented favouring ziprasidone for negative symptoms and EPS.[115] Ziprasidone was effective against depressive symptoms associated with schizophrenia in one study at a dose of 160 mg/day.[116] Significant improvements in multiple cognitive domains have been reported among ziprasidone treated patients in a variety of treatment contexts.[117] Such changes appear to be unrelated to improvements in other symptoms of schizophrenia. Ziprasidone treatment has been associated with significant improvement in quality of life measures in one post hoc data analysis.[118] Further investigation of the effect of ziprasidone on health related quality of life and similar outcomes are warranted. Ziprasidone treatment of schizophrenia appears to be cost-effective relative to no treatment.[119] Further cost-benefit studies are needed.

The dose range of ziprasidone for acute treatment appears to be between 80 and 160 mg/day, higher doses within this range may be more effective (Table 6.2.5.5). Doses greater than 120 mg/day appear to be required to achieve >60 per cent dopamine D2 receptor blockade,[120] the D2 receptor occupancy threshold that appears to coincide with efficacy against positive symptoms, as presented earlier. The medication is usually given twice daily, although some may take the medication once daily at night time. A titration of the total daily dose into the recommended range is required after initiating the medication. One critical aspect of medication administration for ziprasidone is the requirement that the medication be taken with food. There appear to be profound differences in bioavailability at equivalent doses between the fed and unfed state.[121] A full meal, as opposed to a light snack, appears to be required. Therefore, patients are encouraged to take their medication with meals.

A short-acting intramuscular formulation of ziprasidone has been developed which should be useful in situations where more rapid action is needed. This formulation is available in two doses (10 and 20 mg), the preferred dosage being 20 mg due to significantly greater reduction in agitation relative to lower dose.[122] The use of the short-acting injectable form can facilitate a transition to oral medication, and may reduce that time required to titrate the daily dose of ziprasidone to one that is likely to be effective.

Ziprasidone appears to be well tolerated. Treatment-emergent EPS burden is low.[112,113] Initial problems with somnolence or behavioural activation are usually self limited, although temporary use of clonazepam or other benzodiazepine at low doses may improve tolerability, especially during the titration phase, should the latter occur. Importantly, data from both short- and long-term studies indicated that ziprasidone is not associated with clinically significant changes in weight, glycaemic measures, or markers of lipid homeostasis.[123]

Ziprasidone can result in partial blockade of the slow potassium rectifier current in the cardiac conduction system, which may result in prolongation of the QTc interval (discussed in greater detail below).[124] On the other hand, there is only one case report of ziprasidone induced *torsades de pointes*, the risk of which is believed to be increased if the QTc interval is >500 msec.[125] There have also been no reported deaths in the context of overdose with the medication. Under routine circumstances, screening electrocardiograms are not required. Nevertheless, caution may be warranted for individuals who are at risk for significant prolongation of the QTc, including patients who take medications other than ziprasidone that prolong the QTc. Concomitant use of CYP-450 3A4 inhibitors does not appear to pose a significant risk.[126]

In summary, ziprasidone appears to be a useful additional atypical antipsychotic agent because of its favourable side-effect profile, including no weight gain—a major problem with olanzapine and clozapine-and no prolactin elevation, which is a less serious side-effect of risperidone. Patients should be instructed to take the medication with food. Ziprasidone treatment may result in an increase in the QTc interval; however, in a great majority of cases, this is not clinically significant.

(vi) Aripiprazole

Among the atypical antipsychotics, aripiprazole is pharmacologically unique in that it combines partial D2 receptor agonism with high potency 5-HT2A antagonism. Because it is a partial D2 receptor agonist, it binds to the receptor with full affinity, but exerts only a fraction of the intrinsic activity at that site that would be expected of endogenous dopamine. As such, in states of relative dopamine excess, as is believed to be the case in the ventral striatum among schizophrenic patients who experience positive symptoms, aripiprazole is believed to exert relative antagonist activity at D2 receptors.[127] Conversely, it is believed to act primarily as an agonist in cases of relative hypo-dopaminergia, as may be the case in the prefrontal cortex in patients with schizophrenia.[22] For this reason, aripiprazole and other D2 partial agonists in development are sometimes referred to as 'dopamine stabilizers.' Aripiprazole also functions as a potent 5-HT1A partial agonist.[128]

The efficacy of aripiprazole in the treatment of acute schizophrenia at doses ranging between 10 and 30 mg (taken once daily) was established on the basis of four short-term randomized controlled studies.[129] Relative to placebo, efficacy against negative symptoms was also demonstrated.[130] Long term superiority of aripiprazole (vs. placebo) for relapse over 26 weeks[131] and medication compliance and symptom response (vs. haloperidol) for up to 52 weeks has also been established.[133] One study reported on the effectiveness of flexibly dosed aripiprazole (15–30 mg daily) among patients with schizophrenia with a history of resistance to treatment with olanzapine or risperidone.[133] The utility of aripiprazole in the setting of well defined treatment refractory schizophrenia requires further systematic investigation.

The overall effectivenss of aripiprazole has been evaluated in two recent studies. One study reported the effectivenss of flexibly dosed aripiprazole over 8 weeks of treatment (53 per cent response rate at mean endpoint dose = 19.9 mg/day) among a cohort of patients with chronic schizophrenia and schizoaffective disorder under routine treatment conditions in a community healthcare setting.[134] The second study documented comparable effectiveness with olanzapine over 52 weeks of treatment, with more favourable effects for aripiprazole for several metabolic adverse effects.[135] As is the case with other atypical antipsychotic drugs, early evidence indicates that aripiprazole may also have beneficial effects on neurocognitive performance in patients with schizophrenia at recommended doses.[136] Further study of the effects of this medication on cognition is indicated. Clinically relevant improvement in quality of life has been documented in one study.[137] More cost-outcome studies of aripiprazole are needed.

Treatment with aripiprazole is usually initiated with 10 to 15 mg daily, although some patients may not be able to tolerate these doses due to agitation, nausea or vomiting. The dose can be increased up to 30 mg if needed, and tolerated. An oral solution form is also available. Aripiprazole is also available in a soluble

wafer as well as an acute intramuscular form. The acute injectable form appears to be effective in the dosage range of 5.25 to 15 mg.[138] The recommended dose is 9.75 mg.

Aripiprazole is generally well tolerated, with an adverse effect profile similar to placebo in short-term studies involving patients with acute schizophrenia and in longer-term studies of chronic, stable patients.[129] As is the case with all atypical antipsychotic drugs, the EPS burden is lower than that of typical neuroleptics. Some patients, however, may encounter this effect if the dose is started too high or if the titration is too aggressive. Aripiprazole treatment does not appear to significantly increase, and may cause a slight decrease, in prolactin levels.[25] Importantly, short- and longer-term studies indicated that, similar to ziprasidone, aripiprazole is not associated with a high risk of significant changes in weight, glycaemic measures, or markers of lipid metabolism.[123]

In summary, aripiprazole appears to be effective as an acute and long-term maintenance treatment for schizophrenia and related psychotic disorders at recommended doses, though some patients may require higher doses. Aripiprazole was initiated in most studies at doses of 10 to 15 mg once daily; however, some patients may require a slower titration following a lower starting dose. This medication is available in many dosing forms, all of which appear to be very well tolerated. Important benefits from a tolerability viewpoint include very low rates of prolactin elevation, and low risk of weight gain and metabolic adverse effects.

(vii) Paliperidone

Paliperidone is the most recent antipsychotic drug to gain approval for use in the US. It is the 9-OH metabolite of risperidone, which has a longer elimination half-life than the parent compound, as reviewed above. Additionally, paliperidone, which is pharmacologically similar with regard to receptor occupancy profile to risperidone, is available commercially in an extended release form.

The short-term efficacy of paliperidone has been established on the basis of three randomized, placebo controlled studies, two of which have been published,[139,140] that investigated the clinical efficacy of 5 fixed doses (3, 6, 9, 12, and 15 mg) given once daily relative to placebo. In each of these studies, all doses of paliperidone were superior to placebo for reducing global psychopathology and positive symptoms, as well as negative symptoms, anxious/depressive symptoms associated with schizophrenia, and hostility/excitement. In addition, all doses of paliperidone were superior to placebo for improving measures of functional capacity. Paliperidone has not been investigated in the context of treatment refractory schizophrenia. Paliperidone appears to be effective for the prevention of relapse on the basis of one published study.[141]

The recommended starting dose of paliperidone in its extended release form is 6 mg, given once daily. Even though there is a suggestion of greater improvement in terms of symptom reduction from the paliperidone registration studies at higher doses, the adverse effect burden may also be greater (discussed below). Doses may be upwardly adjusted at 3 mg/day increments, up to 12 mg daily. Investigatons of doses greater than 6 mg daily for patients who do not respond adequately have not been performed. A long acting injectable form of paliperidone is currently in development.

Pooled analysis of data from the three short-term, acute phase studies indicate that paliperidone appears to be well tolerated, and that the recommended starting dose (6 mg once daily) was associated with a placebo-like overall adverse effect profile.[142] At doses higher than 6 mg, there appeared to be an increase in the reported incidence of EPS, though not to the degree at any of the doses tested that would be expected with typical neuroleptic treatment. Elevations in prolactin levels appear to be consistent with those observed with risperidone treatment, and appear to be greater in magnitude at higher doses. This effect appears to be especially pronounced among female patients. There were no significant changes from baseline in weight or measures of lipid or glucose handling. Data from long term investigations will provide a more comprehensive picture of paliperidone's tolerability profile.

In summary, paliperidone, the newest atypical antipsychotic drug, appears to be safe and effective for both short- and long-term treatment of schizophrenia. The EPS and prolactinemic adverse effect burden may resemble that of risperidone, but this notion requires prospective investigation. Paliperidone in its extended release form can be started at a clinically effective dose. A long-acting injectable form is currently in development.

(viii) Amisulpride

The efficacy of amisulpride for the treatment of positive symptoms has been established over a wide dosage range (200 to 1200 mg daily) in treatment studies of up to 12 months duration.[143] In general, it appears that higher doses (above 400 mg/day) are effective for treating patients with predominately positive symptoms, although efficacy against negative symptoms has also been demonstrated in this dosing range.[144,145] Low-dose amisulpride (≤ 300 mg/day) has been shown to be effective in treating negative symptoms in schizophrenics with predominantly negative symptoms.[146–148] Evaluation of the effect of amisulpride in patients with minimal extrapyramidal side-effects and positive symptoms suggests amisulpride is able to improve primary negative symptoms, even in patients with deficit syndrome schizophrenia.[146–148] At both dose ranges, amisulpride produces minimal extrapyramidal side-effects, but may result in increased prolactin levels. Amisulpride has been directly compared with haloperidol, and with both risperidone and olanzapine. In general, amisulpride appears to be as clinically effective as all three drugs for treating positive symptoms. Improvement in negative symptoms is superior to haloperidol and appear equivalent to olanzapine and risperidone. Improvement in depressive symptoms related to schizophrenia were also equivalent between amisulpride and olanzapine; however, in a meta-analysis of three studies, amisulpride was shown to be superior to high dose risperidone (8 mg daily) and haloperidol.[149] It is unknown if it is effective in neuroleptic-resistant patients. Because its pharmacology is quite distinct from that of the 5-HT_{2A}-based receptor antagonists previously discussed, amisulpride may be useful in patients who fail to tolerate that class of drugs. Amisulpride has also been demonstrated as being superior to haloperidol on quality of life measures and global functioning.[150] Pharmacoeconomic analyses indicate that amisulpride has a beneficial cost-outcome profile.[151]

(c) Iloperidone and asenapine

Clinical trials are currently taking place with both of these atypical agent to determine its efficacy and side-effect profile compared with typical and other atypical antipsychotic drugs. Like most other atypical antipsychotic drugs, both asenapine and iloperidone combine potent 5-HT2A antagonism with less potent D2 receptor antagonism. Asenapine is currently undergoing investigation in phase 3

clinical trials. Short-term, acute phase efficacy of iloperidone for symptom reduction relative to placebo has been demonstrated at daily doses of 20 to 24 mg daily, with less certain effects at lower doses.[152] Long-term investigations thus far indicate a low incidence of EPS, lack of effect on prolactin release, and minimal effect on body weight.[152] It has the potential to be made into a long-acting form, which would be of great value.

Pharmacokinetics, metabolism, and drug interactions

(a) Typical neuroleptics

The typical neuroleptics are well absorbed when administered orally or parenterally. Intramuscular injection leads to more rapid and higher plasma levels. Peak plasma levels are reached in 30 min after intramuscular injection and 1 to 4 h after oral injection. Steady state is achieved in 3 to 5 days. The half-life for elimination is in the range of 10 to 30 h. Substantial amounts of the antipsychotics are stored in lipids, including in the brain. There is controversy about how long these drugs persist in the system after discontinuation. By the criterion of elevations of plasma prolactin levels, the concentrations are too low to be biologically active within 48 h after discontinuing oral medication. On the other hand, some rodent and human positron emission tomography studies suggest that long-acting forms of haloperidol or fluphenazaine may persist for 1 to 3 months. Metabolism of the typical and atypical antipsychotic drugs occurs in the liver for the most part, via conjugation with glucuronic acid, hydroxylation, oxidation, demethylation and sulphoxide formation. Much of this metabolism occurs via the hepatic cytochrome (CYP)-450 enzymes, particularly the 2D6 and 3A4 sub-families for most drugs. Some metabolites have significant biological activity, for example mesoridazine, and 7-hydroxyloxapine. Dosing of typical neuroleptic medications are determined by clinical effects, less by pharmacokinetic factors.

Pharmacokinetic drug-drug interactions at the level of protein binding are expected to be minimal, even though most typical neuroleptics are tightly bound to plasma proteins. Even so, appropriate therapeutic monitoring of drugs that are also tightly bound to plasma proteins but have a narrow therapeutic index (e.g. warfarin, digoxin, phenytoin) when used in conjunction with typical neuroleptics is warranted. Interactions at the level of the CYP-450 system are also thought to be minimal for most agents. Because smoking is so common among patients with schizophrenia and because smoking can be associated with potent induction of CYP-450 1A2 isoenzymatic activity, dosage adjustments may be needed for selected antipsychotic drugs during any changes in smoking status. Other combinations with typical neuroleptics may be worth avoiding for other reasons, such as increased central nervous system affects (e.g. anxiolytics, other central nervous system depressants, anticholinergics, certain antihypertensive drugs), increased EPS (e.g. metoclopramide, D2 blocking anti-nausea drugs, caffeine), impaired cardiac conduction (certain drugs combined with typical neuroleptics known to prolong the QTc interval), and neurotoxicity (lithium), especially among individuals who are more advanced in age.

(b) Atypical antipsychotic drugs

(i) Clozapine

There are wide variations in the pharmacokinetics of clozapine in patients. The average half-life is 6 to 12 h. Plasma concentrations are higher in Chinese patients than in Caucasian patients, in non-smokers than smokers, and in females than males. The bioavailability is not affected by food intake, metabolism occurs mainly in the liver. The chief metabolite is *N*-desmethylclozapine, which has some biological activity. Clozapine is metabolized by CYP1A2, and several potential drug-drug interactions are thus possible. When agents that induce CYP-1A2 are prescribed or ingested, close monitoring of patients for a worsening of symptoms is warranted. Plasma levels of clozapine of approximately 350 ng/ml are more often associated with good response than lower levels,[153] and should be checked in such cases. Upward adjustment of the clozapine daily dosage will typically correct the problem. On the other hand, if a CYP1A2 inducer is discontinued or a potent inhibitor is added, this may result in a rise in clozapine concentration, and an increase in adverse effect risk.[154] Caution may also be warranted for drugs that are potent inhibitors of CYP 2C19 and CYP3A4.[155] In addition, caution is warranted when considering concomitant use of drugs which can also cause bone marrow suppression (e.g. carbamazepine) or precipitously drop seizure threshold.

(ii) Risperidone

Risperidone is well absorbed from the gut and is extensively metabolized in the liver by CYP2D6 to 9-hydroxyrisperidone in approximately 92 to 94 per cent of Caucasians.[156] Thus, 9-0H risperidone is an active species in the majority of patients. About 6 to 8 per cent of Caucasians and a small proportion of Asians have a polymorphism of the CYP2D6 gene, which leads to poor metabolism of risperidone. For poor metabolizers of risperidone, the active moiety is mainly the parent compound. The half-life of the 9-hydroxy metabolite is about 21 h whereas the half-life of risperidone is about 3 h. Thus, risperidone can be used on a once-a-day schedule for normal metabolizers whereas multiple doses are needed for those who are poor metabolizers. Risperidone should be titrated from 2 to 5 mg/day over at least a 3-day period to minimize hypotensive and neuro muscular side-effects. Drugs known to induce or inhibit CYP2D6 and 3A4 may alter plasma levels of risperidone; thus, close monitoring is advised when such agents are added to ongoing risperidone treatment.

(iii) Olanzapine

Olanzapine has a half-life of 24 to 30 h, which indicates that single daily administration is adequate.[92] The metabolic pathways of olanzapine involves CYP2D6, CYP1A2 and flavin-containing mono-oxygenases, as well as *N*-glucuronidation. It has a low potential for drug-drug interactions and requires extremely high concentrations not likely to be achieved under clinical conditions to inhibit cytochrome P-450 systems. Plasma levels of approximately 9.3 mg/ml have been reported to predict better clinical response to olanzapine in inpatients with an acute exacerbation.[157] Drugs that are known inducers or inhibitors of CYP1A2 may significantly affect plasma levels of olanzapine and alter its clinical effects at a given dosage; thus, active monitoring of symptoms and adverse effects is indicated if such agents are added. As is the case with clozapine, gender and smoking status may influence olanzapine levels leading to adjustment in dosage.[153]

(iv) Quetiapine

Quetiapine is well absorbed and is approximately 83 per cent protein bound.[158] Quetiapine is absorbed better after eating.[158] It has a half-life of 6 h. It is metabolized in the liver by CYP3A4 to inactive metabolites. Quetiapine has significant interactions with several inducers and inhibitors of CYP3A4. Co-administration with

these agents may require dosage adjustment. Thioridazine may also significantly increase the clearance of quetiapine,[159] thus necessitating dosage adjustment. Despite the short half-life, a clinical trial compared three dosing regimens (450 mg/day given in two or three divided doses, and 50 mg/day given twice daily). Both of the higher-dose groups were superior to the low-dose group and there were no differences between the two high-dose schedules. Once daily dosing, which is also a common dosing strategy for quetiapine, is also supported in the literature.[160] The feasibility of such a dosing schedule, which does not seem to be predicted by peripheral pharmacokinetic parameters, is possible because quetiapine appears to interact centrally with both D2 and 5-HT2A receptors much longer than its 6 h elimination half-life.[161]

(v) Ziprasidone

Ziprasidone has a half-life of 4 to 10 h. Twice-daily administration is possible despite this relatively short half-life. Clinically, many patients are prescribed this medication only once daily. Regardless, ziprasidone should always be taken after eating in order to facilitate absorption. About two-thirds of ziprasidone is metabolized by aldehyde oxidase into inactive metabolites. The remainder is metabolized by CYP3A4 and CYP1A2 into inactive metabolites. At the current time, there are no known drug interactions with ziprasidone at the level of aldehyde oxidase, since enzymatic activity does not appear to be altered by coadministered drugs. Although CYP3A4 appears to play only a minor role in the metabolism of ziprasidone, potent inhibitors or inducers of CYP3A4 may significantly alter plasma concentrations of ziprasidone,[162] and may thus necessitate an adjustment in dosage. The use of concomitant medications that may prolong the QTc interval should be avoided. Ziprasidone is contraindicated for patients with a history of known QT prolongation, recent acute myocardial infarction, or uncompromised heart failure.

(vi) Aripiprazole

Aripiprazole is well absorbed from the gut, and has an elimination half-life of 75 hours. It is metabolized primarily by CYP3A4 and 2D6 isoenzymes into an active metabolite, dehydro-aripiprazole, which has a half-life of 94 hours. This pharmacokinetic pattern supports once daily dosing. Because aripiprazole is metabolized by CYP3A4 and 2D6, known inhibitors or inducers of these isoenzymes may result in increased or decreased clearance of aripiprazole and dehydro-aripiprazole.[163]

(vi) Paliperidone, Iloperidone, and Amisulpride

Paliperidone is currently marketed in the US and abroad only in an osmotically controlled extended release formulation, which results in steady release of active drug over a 24 hr period. Hepatic metabolism is not considered a major route of clearance. Paliperidone is converted into metabolites that are not believed to contribute significantly to its overall pharmacological activity. Few significant drug-drug interactions at the level of the CYP450 system are therefore anticipated. Even so, the plasma concentration of paliperidone may be altered by drug interactions at CYP3A4.[164]

Iloperidone has a half-life of 12 to 15 h. Its absorption is not affected by food. It should be titrated slowly because of orthostatic hypotension, and close monitoring is warranted when it is combined with antihypertensive drugs or drugs that are associated with orthostatic effects. The optimal dose has not yet been established but is likely to be in the 5 to 10 mg/day range. Amisulpride has a half-life of 10 to 15 h. It is well tolerated. As yet, there are no known drug interactions. More information regarding the metabolic handling and potential for drug-drug interactions for both of these medications is anticipated as they continue to be further developed.

Side-effects

Typical neuroleptics

The adverse effects that are most routinely concerning for antipsychotic drug treatment are extrapyramidal adverse effects (EPS), especially for typical neuroleptic mediations. For typical neuroleptics, high-potency drugs such as haloperidol and fluphenazine are more likely to produce EPS than low-potency agents such as chlorpromazine and thioridazine. The latter may have lower potential for extrapyramidal side-effects than other typical neuroleptics because of its relatively higher affinity for muscarinic receptors. Atypical antipsychotic drugs are less likely to cause EPS during acute and long term tretment. There are a wide range of extrapyramidal side-effects produced by the typical neuroleptics, including dystonic reactions when first administered, akathisia during the first 2–3 weeks, parkinsonism during the first several weeks with variable persistence, neuroleptic malignant syndrome at any time point but usually in the initial weeks, and tardive dyskinesia.

Dystonic reactions due to neuroleptic drugs can be treated with parenteral anticholinergic agents or diphenhydramine, an antihistamine with some anticholinergic properties. The use of anticholinergic and other agents to manage parkinsonism due to typical neuroleptic drugs will be discussed subsequently.

Akathisia may be the most common of the EPS effects, occurring in up to 70 per cent of patients treated long term with haloperidol.[165] The term refers to a subjective uncomfortable experience of motor restlessness which is relieved by movement. As such, patients will complain of discomfort, and manifest increases in psychomotor behaviour. These symptoms can be so distressing as to increase the risk of agitation or even suicidal behaviours.[166] Although patient age does not seem to influence risk of developing akathisia, women are believed to be at higher risk. Accurate diagnosis of this condition is necessary in order to prevent inadvertent increases in neuroleptic dose from a belief that the patient's discomfort from akathisia is due instead to worsening psychosis. This effect may be managed by reduction in dosage or switching medications to an atypical antipsychotic drug or a drug that is less likely to cause akathisia. When these strategies are not feasible, the symptoms may respond to anticholinergic medications, usually within 3–7 days. Other options include low doses of benzodiazepines or beta-adrenergic blockers, assuming no contraindications to either.

Parkinsonism caused by antipsychotic drugs resembles idiopathic parkinsonism. Diagnostically, severe neuroleptic induced parkinsonism may resemble depression or negative symptoms of schizophrenia; however, the associated motor signs and time course of symptoms in relation to starting antipsychotic treatment distinguish the former. Like akathisia, the onset and severity of antipsychotic induced parkinsonism is related to medication

dosage; thus, a lowering of the dose or switching to a medication that is less likely to cause this effect may provide significant relief, or ameliorate the parkinsonian signs and symptoms altogether. When this is not feasible, anticholinergic medications may provide relief, typically within 3–7 days. The response to anticholinergic medication is quite variable, however.

Tardive dyskinesia emerges at various rates depending upon age, sex, and diagnosis.[167,168] The rate in younger patients is between 3 and 5 per cent per year. It is higher in bipolar than schizophrenic patients and much higher in people above the age of 60. It is related to dose and will be less likely with lower doses of typical neuroleptics. Tardive dyskinesia is ordinarily reversible, although irreversible and/or extremely severe and rarely life-threatening forms can occur. The best way to minimize its occurrence is to use an atypical antipsychotic drug in lieu of a typical agent, since these drugs as a class are associated with a much lower risk of tardive dyskinesia.[167] Patients with mood disorders should generally not receive maintenance treatment with typical antipsychotic drugs unless mood stabilizers alone prove insufficient because they are at greater risk for tardive dyskinesia. There are no definitive treatments for tardive dyskinesia. Generally, the best strategy is prevention through the use of atypical antipsychotic drugs, and periodic screening with a structured assessment tool such as the Abnormal Involuntary Movement Scale (AIMS). There is some suggestion in the literature that continuation of antipsychotic treatment does not worsen tardive dyskinesia, and may eventually result in a stabilization and improvement of tardive symptoms. Switching to clozapine appears to be helpful, although such an effect is not invariable.

Neuroleptic malignant syndrome is a rare life-threatening side-effect related to an apparent compromise of the neuromuscular and sympathetic nervous systems.[169] It usually occurs at the initiation of treatment with a high-potency agent but may occur with any of the typical (or atypical agents) at any point. Immediate discontinuation of the medication is essential. The condition is characterized by muscle rigidity, breakdown of muscle fibres leading to large increases in plasma creatine kinase activity, fever, autonomic instability, changing levels of consciousness, and sometimes death. It may be treated by discontinuing all antipsychotic drug treatment, applying external hypothermia, supporting blood pressure, and administering a direct-acting dopamine agonist such as bromocriptine or pergolide, and dantrolene sodium, which blocks the release of intracellular stored calcium ions. After its successful treatment, an atypical antipsychotic should be used even though these agents, including clozapine, may also induce neuroleptic malignant syndrome.

The typical neuroleptic drugs produce a wide variety of **other side-effects**, including weight gain, seizures (especially pimozide), sedation, hypotension, elevated liver enzymes, retinitis pigmentosa (thioridazine), orthostatic hypotension, prolongation of the QTc interval (low potency phenothiazines, pimozide) and anticholinergic effects (mesoridazine, chlorpromazine, thioridazine). All the typical neuroleptic drugs produce marked increases in serum prolactin levels, with the increases being greater in females than males.[25] Prolactin elevations may affect sexual function in both males and females, with difficulty achieving erection or orgasm among the most common side-effects.[25]

Atypical antipsychotic drugs

(a) Clozapine

(i) Agranulocytosis

It has now been reliably established that clozapine produces agranulocytosis in slightly less than 1 per 100 patients.[170, 171] This is 15 to 30 times the rate associated with the phenothiazines and possibly higher than that for the butyrophenones. The peak of agranulocytosis with clozapine occurs between 4 and 18 weeks, and then falls off sharply. Weekly monitoring of the white cell or absolute neutrophil count is required for 26 weeks in most countries, with the frequency decreasing to biweekly or monthly thereafter, sometimes on a voluntary basis. In the US, monthly monitoring is required assuming no hematological abnormalities after one year of treatment. The cost-effectiveness of monitoring after a year has not been studied but it is probably in the range that would lead to its abandonment by current standards. With monitoring, agranulocytosis can usually be detected before infection sets in or becomes overwhelming. Discontinuation of clozapine, beginning treatment with colony cell stimulating factors, and the usual procedures for treating an infection are usually effective in restoring the white cell line.

(ii) Other side-effects

Clozapine produces a wide range of side-effects.[171] These can generally be managed by dose adjustment and concomitant medications. Clozapine produces hypotension because of its potent α_1-adrenoceptor antagonism and must be slowly titrated in most patients. Low-dose glucocorticoid treatment may be helpful in some patients with severe hypotension. Clozapine rarely if ever produces significant extrapyramidal side-effects, although some cases of akathisia and neuroleptic malignant syndrome have been reported.

Major motor seizures are another important side-effect of clozapine. They are dose related, with the incidence being about 2 per cent in patients at low doses and 6 per cent at doses greater than 600 mg/day. They are sometimes preceded by myoclonic jerks. Valproic acid and dose reductions are usually effective in preventing the progression of myoclonic jerks or treating major motor seizures. Other anticonvulsants can be combined with clozapine if needed, though caution would be clearly warranted with the use of carbamazepine due to its potential for bone marrow suppression.

Hypersalivation is another side-effect. It usually responds to anticholinergic therapy or to clonidine. Exacerbation of obsessive-compulsive symptoms has been reported with clozapine. Augmentation with an SSRI or lithium carbonate is usually effective.

Weight gain is a frequent side-effect of clozapine, with about 30 per cent of patients gaining more than 7 per cent of body weight.[171] Diet and exercise are useful in minimizing this effect. A related problem is the emergence of insulin resistance or type II diabetes, or exacerbation of existing diabetes, with or without atherogenic changes in serum lipid profile. There have also been reports of diabetic ketoacidosis that emerged in the context of clozapine treatment. Of the atypical antipsychotic drugs, clozapine and olanzapine are associated with the highest risk for clinically significant weight gain, as well as abnormalities in glycaemic control and lipid homeostasis.[123]

Somnolence, tachycardia, hypertension, constipation and stuttering are also produced by clozapine. Tachycardia is treated only when the pulse is greater than 100 beats/minute. β-Blockers are effective to reduce the heart rate, but may also result in synergism of hypotensive effects.[171]

There have been reports of clozapine-associated myocarditis and cardiomyopathy.[172] The presence of eosinophilia accompanied by cardiotoxic signs such as tachycardia, fatigue, orthostasis, or respiratory problems (many of which are adverse effects of clozapine) should alert the clinician to the possibility of myocarditis and the need for medical evaluation.

Finally, treatment with clozapine may not uncommonly result in an asymptomatic mild elevation in hepatic transaminase levels; however, there have also been reports of hepatotoxicity in the setting of clozapine treatment. Polypharmacy appears to be a risk factor. Cases of fulminant hepatotoxicity leading to liver failure are rare.

(iii) Risperidone

Risperidone is associated with moderate weight gain, comparable to that of typical neuroleptic drugs in most cases, and less than that of clozapine and olanzapine.[123,173] Risperidone also produces some postural hypotension because of its α_1-adrenoceptor blocking properties. Risperidone produces greater increases in serum prolactin secretion than any of the other atypical antipsychotic drugs.[25] The increases appear to be at least comparable to those of typical neuroleptics.[173] At higher doses, particularly above 6 mg daily in most adults, the incidence of EPS also increases,[77] though typically not to the degree observed when using typical neuroleptic drugs in clinical practice. Risperidone, like clozapine and other agents of this type, can sometimes exacerbate or induce symptoms of obsessive-compulsive disorder and tics, probably due to its antiserotonergic properties. This can be counteracted in some patients by the addition of an SSRI. Risperidone is not associated with agranulocytosis or increased risk of seizures. Because of its low affinity for muscarinic receptors, risperidone treatment is not associated with significant anticholinergic effects.

(iv) Olanzapine

Olanzapine also produces dose-dependent extrapyramidal side-effects, including some dystonic reactions in patients with schizophrenia, but these are less frequent and severe than those produced by typical neuroleptic drugs or risperidone.[173] Olanzapine is less well tolerated than clozapine in patients with Parkinson's disease. Olanzapine, like other atypical antipsychotic drugs, is associated with a lower risk of tardive dyskinesia than typical neuroleptics.

The major side-effect of olanzapine is weight gain.[173] Large weight gains due to increased appetite occur in 10 to 15 per cent of olanzapine-treated patients during the first 6 months of treatment. Another 20 to 35 per cent gain between 7 and 10 per cent of body weight. These gains tend to become permanent for as long as patients continue the medication. Like clozapine, olanzapine is also associated with higher risk of insulin resistance, glycaemic changes, and development of atherogenic changes in lipid profile.[123] Cases of diabetic ketoacidosis associated with olanzapine treatment have been reported.

Olanzapine is also associated with some increase in liver enzymes, orthostatic hypotension, anticholinergic side-effects, and sedation. Many of these adverse effects are time limited and reduce in intensity or resolve over the first few weeks of treatment with continuous use. Olanzapine produces transient increases in serum prolactin levels, which are smaller in magnitude than those produced by typical neuroleptic drugs or risperidone.[25, 173]

Olanzapine, like other agents of this type, can occasionally exacerbate or induce symptoms of obsessive-compulsive disorder and tics, probably due to its antiserotonergic properties. This can be counteracted in some patients by the addition of an SSRI. Olanzapine is not associated with agranulocytosis or increased risk of seizures.

(v) Quetiapine

Quetiapine appears to have fewer extrapyramidal side-effects than either risperidone or olanzapine.[158,173] Quetiapine is tolerated in patients with Parkinson's disease to a much greater extent than risperidone or olanzapine. The incidence of extrapyramidal side-effects with quetiapine in schizophrenic patients appears to be comparable to placebo. The major side-effects with quetiapine are headache, agitation, dry mouth, dizziness, weight gain, and postural hypotension.[173]

With regard to weight gain and other metabolic effects, quetiapine treatment appears to confer moderate risk—similar to that of risperidone, but less than that associated with clozapine or olanzapine treatment.[123] Far less is known about the long term effects of quetiapine on markers of glycaemic and lipid homeostasis. Nevertheless, clinically significant changes in serum lipids have been reported.

Decreased serum thyroid hormone levels, increased hepatic transaminases and elevated serum lipids have been reported. Decreases in total and free thyroxine, when they occur, are mild, non-progressive, and are not believed to be clinically signficiant. The effect may be dose dependent. Similar to clozapine, asymptomatic elevations in hepatic transaminases may be encountered early in the course of treatment, followed by a return to baseline values. Animal studies suggest an increased risk of cataracts.[173] Periodic ophthalmological screening for lenticular opacities is recommended by the manufacturer, though no causal relationship between the use of quetiapine and the development of cataracts has been demonstrated to date.

(vi) Ziprasidone

Ziprasidone does not increase serum prolactin levels and is virtually devoid of extrapyramidal side-effects, weight gain, and changes in markers of glucose handling and lipid metabolism. Its major side-effects are nasal congestion and somnolence,[173] the latter of which is usually transient. There has been some concern of cardiovascular side-effects, for example increased QTc interval; however, perusal of the available data does not reveal a significant problem in this regard. However, caution is warranted when considering the coadministration of ziprasidone with other drugs that are known to prolong the QTc interval, since ziprasidone has been associated with a significant increase in the QTc interval of 16.6 msec, which was greater than that of other atypical antipsychotics and haloperidol, but less than thioridazine.[174] Screening for electrolyte abnormalities and cardiac disease (including recent myocardial infarction, congestive heart failure symptoms and arrythmias with or without syncope) may be indicated prior to starting ziprasidone.

(vii) Aripiprazole and paliperidone

Aripiprazole is well tolerated, and does not appear to routinely cause EPS or hyperprolactinaemic changes at recommended dosages. This also appears to be the case for higher than recommended doses. Aripiprazole is also not associated with clinically significant increases in weight, or changes in markers of glucose handling or lipid homeostasis. Both aripiprazole and ziprasidone are therefore believed to be the atypical antipsychotic drugs with the most advantageous metabolic risk profile.

Paliperidone in its extended release form also appears to be well tolerated during short term, acute phase treatment. In these studies, the most common side effect was tachycardia. Rates of discontinuation due to adverse effect burden were also very low. The risk of hyperprolactinaemic changes with paliperidone appears to resemble those of risperidone, although no head-to-head comparisons have been carried out. The changes in prolactin levels may be dose related. The EPS burden associated with paliperidone during the short term studies was low for the 6 mg dose; however, at higher doses, the incidence of EPS appears to be higher. Measures of weight and metabolic effects during 6 week treatment with paliperidone showed no significant changes from baseline. Similar results were found for paliperidone during medium-term treatment.[141] Future long term studies will add greatly to our understanding of paliperidone's adverse effect profile.

Indications and contraindications

The main indication for the antipsychotic drugs is the treatment of all phases of schizophrenia, including acute, florid symptoms of psychosis, prevention of relapse, and deficit symptoms. Important other uses include the psychotic phase and prophylaxis of mania, depression with psychotic features, the psychosis, agitation, and aggression of various dementias, the treatment of psychoses due to l-dopa or other dopamine agonists in Parkinson's disease, Tourette's syndrome, treatment-resistant obsessive-compulsive disorder, self-injurious behaviour, porphyria, antiemesis, intractable hiccoughs, and as antipruritics. Some current research has suggested that the antipsychotic drugs may be of use to prevent the onset of schizophrenia by administering them to individuals who are in the prodromal phase of the illness. The atypical antipsychotics may be effective for augmenting antidepressants in patients with treatment-resistant non-psychotic depression, and are being tried on an experimental basis for various character disorders such as borderline, schizoid, and schizotypal personality disorders. Clozapine, which has the lowest incidence of extrapyramidal side-effects of any of the antipsychotic drugs, has some special applications in neurological conditions such as essential tremor and the treatment of the water intoxication syndrome in schizophrenic patients. The uses of the classical antipsychotics such as chlorpromazine and haloperidol have been limited by their side-effects, especially parkinsonism and tardive dyskinesia, a slowly developing, sometimes irreversible series of abnormal involuntary movements involving facial, limb, and girdle muscles. As has been discussed, the atypical antipsychotic drugs such as clozapine, olanzapine, quetiapine, and risperidone, as well as iloperidone and ziprasidone, which are in development, have significant advantages with regard to parkinsonism. Clozapine definitely has a vastly reduced risk of tardive dyskinesia and the other atypical agents most likely have a risk that is less than that of the typical neuroleptic drugs but more than clozapine. Uses in other psychiatric and neurological conditions may be expected to emerge as the safety profile of these agents is better described.

Antiparkinsonian agents

Anticholinergic drugs

Antiparkinsonian medications, including anticholinergic, antihistaminic, benzodiazepines, dopamine agonists, and β-blockers are of importance in the management of extrapyramidal side-effects. They are usually needed with the typical neuroleptic drugs but some patients will require antiparkinsonian treatment with olanzapine, risperidone, or quetiapine. The anticholinergics and the antihistaminics (e.g. diphenhydramine) are used to treat acute dyskinesias and dystonias, pseduoparkinsonian symptoms (tremor, rigidity, bradykinesia, shuffling gait), and akathisia. These agents act centrally in the basal ganglia to block the effects of increased acetylcholine release due to D_2-receptor blockade. The most widely used anticholinergic drugs are benztropine, biperiden, procyclidine, and trihexyphenidyl. Benztropine is given in doses of 1 to 6mg/day usually in divided doses. Biperiden is given in doses of 2 to 16 mg/day in two or three doses. Procyclidine is given in divided doses of 5 to 30 mg/day. Trihexylphenidyl is given in doses of 1 to 15 mg/day, in a single or divided dose.

These agents are competitive antagonists of the five subtypes of muscarinic receptors that have been identified and which are labelled M_1 to M_5. They have minimal antagonist effect at nicotinic cholinergic receptors. Blockade of cholinergic receptors on intrastriatal neurones by these agents restores the cholinergic balance, which is disrupted by blockade of D_2 dopamine receptors by some antipsychotic agents. Other central effects include impairment of various forms of memory. Elderly patients in particular may develop anticholinergic-induced agitation, irritability, disorientation, hallucinations and delirium because of the natural loss of cholinergic neurones with aging.

Side-effects

These agents have some preference for the central nervous system but some peripheral anticholinergic effects are to be expected. Blockade of vagal tone in the heart produces tachycardia. Other adverse effects include decreased bladder function and urinary retention and decreased bowel motility leading to constipation and impaction. Decreased saliva and bronchial secretion contribute to dry mouth and increased dental caries while decreased sweating increases the risk of heat stroke. Blockade of muscarinic receptors in the eye cause pupillary dilation and inhibition of accommodation, leading to photophobia and blurred vision. Rarely, narrow-angle glaucoma may ensue. The muscarinic receptors in the basal ganglia are predominantly M_2 whereas those in the periphery are M_1. The rank order of the anticholinergic drugs for relative selectivity for the M_2 receptor is biperiden, procyclidine, trihexylphenidyl, and benztropine. All these agents can cause dry mouth, blurred vision, urinary retention, constipation, and increased intraocular pressure. They may cause anticholinergic delirium in elderly patients or after taking high doses. Biperiden is less likely to cause peripheral anticholinergic effects. Benztropine, biperiden, and trihexyphenidyl may cause euphoria because of their ability to inhibit dopamine reuptake and may be subject to abuse.

Indications

The anticholinergic drugs or the antihistamine diphenhydramine are given intramuscularly for the treatment of acute dystonic reactions. They are usually effective within minutes and may have to be repeated. It is usually not necessary to prescibe an oral anticholinergic following a dystonic reaction, though some may require their brief use depending on which anipsychotic is prescribed. These agenst should not be given prophylactically unless the patient is at established risk for EPS at the dose of antipsychotic which is being started. If akathisia or parkinsonism develops following treatment with a typical neuroleptic drug, the first consideration should be whether to continue to use the offending agent and drop the dosage or to substitute an atypical antipsychotic drug. If decreasing the dose of antipsychotic drugs does not suffice or is not clinically feasible, substituting an atypical agent is the clearly the recommended choice since it avoids all the unpleasant side-effects of the anticholinergic agents.

Other drugs

Amantadine, which also has antiviral actions, is able to increase the release of dopamine in the basal ganglia, which diminishes the release of acetylcholine. It may improve acute dystonias, akathisia, akinesia, parkinsonism, and tardive dyskinesia. It has also been reported to improve sexual function and decrease weight gain due to neuroleptic drugs. It may cause increased arousal, agitation, and indigestion, however. The usual oral dose is 100 to 400 mg/day.

β-Blockers such as propranolol, atenolol, and pindolol are useful for treating akathisia and tremor. They may cause bradycardia, and particularly immediate-release forms should not be stopped abruptly due to rebound tachycardia.

Benzodiazepines, such as clonazepam, lorazepam, and diazepam, are useful for treating akathisia, acute dystonias, and acute dyskinesias. They can cause drowsiness and lethargy, and have abuse potential.

Conclusions

Antipsychotic drugs are invaluable tools in treating a large variety of patients with schizophrenia and other conditions. Their main benefits are, in fact, to treat psychotic symptoms, but the newer agents in particular may improve negative symptoms, cognition, mood, anxiety, and aggression as well. The evidence for atypical antipsychotic drugs to improve cognition is steadily increasing and this should be one of the driving forces behind the substitution of these agents for the typical antipsychotic drugs. Recent evidence of volumetric increases in cerebral cortical gray matter associated with atypical, but not typical, antipsychotic drugs may be related to improvement in such symptoms.[175] As such, atypical antipsychotic drugs may produce 'disease modifying' rather than just 'symptomatic' effects, a matter that is of considerable current interest.

Antipsychotic drugs are useful as both acute and maintenance treatments to prevent the recurrence of psychotic symptoms. The extrapyramidal side-effects and greater tardive dyskinesia risk of the typical antipsychotics, coupled with their lesser efficacy to improve negative symptoms and cognition suggest that newer agents are preferred. Clozapine, despite its risk of agranulocytosis, is the treatment of choice for patients who fail to respond to other typical or atypical antipsychotic agents. Risperidone, olanzapine, quetiapine, ziprasidone, aripiprazole and paliperidone have somewhat different pharmacologic profiles. It is not clear which of these agents should be tried in a given patient but on going research may clarify that. These agents, and clozapine, appear to differ significantly in their propensity for causing clinically significant changes in weight and markers of metabolic status. Amisulpride has a mechanism of action different from that of the other atypical agents, with some preference for treating negative symptoms. These compounds, as well as others expected to be approved for use in the near future, for example iloperidone and asenapine, will need to be compared with each other to determine if differential indications exist. Side-effect differences among these drugs as well as the availability of long-acting preparations may help clinicians choose among them. Cost-effective analyses currently favour use of the atypical antipsychotic drugs because of better compliance leading to less frequent relapses and shorter hospital stays. They also facilitate retention of work skills and return to work which decreases the indirect costs of illness in patients still young enough to be able to work. As long as the typical antipsychotics remain in use, and for some patients who receive atypical agents, anticholinergic and other antiparkinsonian drugs will continue to be necessary to treat extrapyramidal side-effects.

Because of the compliance problem, which is less with the atypical than the typical antipsychotics, it is important to develop more long acting atypical drugs. Risperidone is currently available in such a form, and paliperidone and olanzapine will also be in the near future. While the current group of atypical antipsychotic drugs is predominantly characterized by relatively more potent 5-HT_{2A} than D_2 receptor antagonism, it is likely that a number of different strategies will emerge for compounds which produce fewer extrapyramidal side-effects than the typical neuroleptics. Because these compounds are so effective in that regard, the real challenge is to develop agents which address other key features of schizophrenia, especially cognitive impairment and negative symptoms, without the side-effect burden of this group of compounds.

Further information

Davis, K.L., Charney, D., Coyle, J., *et al.* (eds.) (2001). *Neuropsychopharmacology: A Fifth Generation of Progress*. New York: Raven Press.

Breier, A., Tran, P.V., Herrera, J.M., *et al.* (2001). *Current Issues in the Pharmacology of Schizophrenia*. Philadelphia: Lippincott Williams & Wilkins.

References

1. Meltzer, H. (1995). The concept of atypical antipsychotics. In: *Advances in the neurobiology of schizophrenia* (eds. J.A. den Boer, H.G.M. Westenberg, and H.M. van Praag), pp. 265–73. Wiley, Chichester.
2. Arndt, J. and Skarsfeldt, T. (1998). Do novel antipsychotics have similar pharmacological characteristics? A review of the evidence. *Neuropsychopharmacology*, **181**, 63–101.
3. Kapur, S. and Seeman, P. (2001). Does fast dissociation from the dopamine D(2) receptor explain the action of atypical antipsychotics? A new hypothesis. *American Journal of Psychiatry*, **158**, 360–9.
4. Dixon, L., Lehman, A., and Levine, J. (1995). Conventional antipsychotic medications for schizophrenia. *Schizophrenia Bulletin*, **21**, 567–78.

5. Fitton, A. and Heel, R. (1990). Clozapine: a review of its pharmacological properties, and therapuetic use in schizophrenia. *Drugs*, **40**, 722–47.
6. Burris, K.D., Molski, T.F., Xu, C., *et al.* (2002). Aripiprazole, a novel antipsychotic, is a highaffi nity partial agonist at human dopamine D2 receptors. *Journal of Pharmacology & Experimental Therapeutics*, **302**(1), 381–9.
7. Szewezak, M., Corbett, R., Rusk, D., *et al.* (1995). The pharmacological profi le of iloperidone, a novel atypical antipsychotic agent. *Journal of Pharmacology and Experimental Therapeutics*, **274**, 1404–13.
8. Alphs, L., Panagides, J., Lancaster, S. (2007). Asenapine in the treatment of negative symptoms of schizophrenia: clinical trial design and rationale. *Psychopharmacology Bulletin*, **40**, 41–53.
9. Schotte, A., Janssen, P.F., Gommeren, W., *et al.* (1996). Risperidone compared with new and reference antipsychotic drugs: in vitro and in vivo receptor binding. *Psychopharmacology*, **124**(1–2), 57–73
10. Meltzer, H. and Fatemi, S. (1996). The role of serotonin in schizophrenia and the mechanism of action of anti–psychotic drugs. In *Serotonergic mechanisms in antipsychotic treatment* (eds. J.M. Kane., H.–J. Moller, and F. Awouters), pp. 77–107. Dekker, New York.
11. Anonymous (1998). Adverse effects of the atypical antipsychotics. Collaborative Working Group on Clinical Trial Evaluations. *Journal of Clinical Psychiatry*, **59**, 17–22.
12. Luft, B., Taylor, D. (2006). A review of atypical antipsychotic drugs versus conventional medication in schizophrenia. *Expert Opinion on Pharmacotherapy*. **7**(13), 1739–48.
13. The Collaborative Working Group (1998). Atypical antipsychotics for treatment of depression in schizophrenia and affective disorders. *Journal of Clinical Psychiatry*, **12**, 41–6.
14. McElroy, S.L., Fry, M., Denicoff, K., *et al.* (1998). Olanzapine in treatment–resistant bipolar disorder. *Affective Disorders*, **49**, 119–22.
15. Calabrese, J., Kimmel, S., Woyshville, M., *et al.* (1996). Clozapine in treatment–refractory mania. *American Journal of Psychiatry*, **153**, 759–64.
16. Keefe, R.S., Silva, S.G., Perkins, D.O., *et al.* (1999). The effects of atypical antipsychotic drugs on neurocognitive impairment in schizophrenia: a review and meta-analysis *Schizohrenia Bulletin*, **25**, 201–22.
17. Woodward, N.D., Purdon, S.E., Meltzer, H.Y., *et al.* (2005). Meta-analysis of neuropsychological change to clozapine, olanzapine, quetiapine, and risperidone in schizophrenia. *International Journal of Neuropsychopharmacology*, **8**, 457–72.
18. Keefe, R., Silva, S., Perkins, D., *et al.* (1999). The effects of atypical antipsychotic drugs on neurocognitive impairment in schizohprenia. *Schizophrenia Bulletin*, **25**, 201–32.
19. Meltzer HY. and Stahl, S.M (1976). The dopamine hypothesis of schizophrenia: A review. *Schizophrenia Bulletin* **2**, 19–76.
20. Davis, K., Kahn, R., Ko, G., *et al.* (1992). Dopamine in schizophrenia: a review and reconceptualization. *American Journal of Psychiatry*, **148**, 1474–86.
21. Meltzer, H. (1999). The role of serotonin in antipsychotic drug action. *Neuropsychopharmacology*, **21**, 106–15.
22. Goldman–Rakic, P. and Selemon, L. (1997). Functional and anatomical aspects of prefrontal pathology in schizophrenia. *Schizophrenia Bulletin*, **23**, 437–58.
23. Abi-Dargham, A., Moore, H. (2003). Prefrontal DA transmission at D1 receptors and the pathology of schizophrenia. *Neuroscientist*, **9**, 404–16.
24. Farde, L., Nordstrom, A., Wiesel, F., *et al.* (1992). Positron emission tomographic analysis of central D1 and D2 dopamine receptor occupancy in patients treated with classical neuroleptics and clozapine. Relation to extrapyramidal side effects. *Archives of General Psychiatry*, **49**, 538–44.
25. Haddad, P.M., Wieck, A. (2004). Antipsychotic-induced hyperprolactinemia: mechanisms, clinical features and management. *Drugs*, **64**, 2291–314.
26. Kramer, M., Last, B., *et al.*, and the D4 Dopamine Antagonist Group (1997). The effects of a selective D4 dopamine receptor antagonism. (L–745, 870) in acutely psychotic inpatients with schizophrenia. *Archives of General Psychiatry*, **54**, 567–72.
27. Richelson, E. (1988). Neuroleptic binding to human brain receptors: relation to clinical effects. *Annals of the New York Academy of Sciences*, **537**, 435–42.
28. Meltzer, H.Y., Matsubara, S., and Lee, M. (1989). Classification of typical and atypical antipsychotic drugs on the basis of dopamine D-1, D-2 and serotonin 2 pKi values. *Journal of Pharmacology and Experimental Therapeutics*, **251**, 238–46.
29. Kuroki, T., Meltzer, H., and Ichikawa, J. (1998). Effects of antipsychotic drugs on extracellular dopamine levels in rat medial prefrontal cortex and nucleus accumbens. *Journal of Pharmacology and Experimental Therapeutics*, **288**, 774–81.
30. Parada, M., Hernande, L., Puig de Parada, M., *et al.* (1997). Selection action of acute systemic clozapine on acetylcholine release in the rat prefrontal cortex by reference to the nucleus accumbens and striatum. *Journal of Pharmacology and Experimental Therapeutics*, **281**, 582–8.
31. Li, X., Perry, K., Wong, D., *et al.* (1998). Olanzapine increases *in vivo* dopamine and norepinephrine release in rat prefrontal cortex, nucleus accumbens and striatum. *Psychopharmacology*, **136**, 153–61.
32. Andersson, J.L., Nomikos, G.G., Marcus, M., *et al.* (1995). Ritanserin potentiates the stimulatory effects of raclopride on neuronal activity and dopamine release selectivity in the mesolimbic dopaminergic system. *Naunyn- Schmiedeberg's Archives of Pharmacology*, **352**, 374–85.
33. Westerink, B.H., Kawahara, Y., De Boer, P., *et al.* (2001). Antipsychotic drugs classified by their effects on the release of dopamine and noradrenaline in the prefrontal cortex and striatum. *European Journal of Pharmacology*, **412**, 127–38.
34. Liegois, J.F., Ichikawa, J., Meltzer, H.Y. (2002). 5HT2A receptor antagonism potentiates haloperidol-induced dopamine release in rat medial prefrontal cortex and inhibits that in the nucleus accumbens in a dose-dependent manner. *Brain Research*, **947**, 1547–65.
35. Armsten, A.F. (2004). Adrenergic targets for the treatment of cognitive deficits in schizophrenia. *Psychopharmacology (Berl)*, **174**, 25–31.
36. Poyurovsky, M., Koren, D., Gonopolsky, I., *et al.* (2003). Effects of the 5-HT2 antagonist mianserin on cognitive dysfunction in chronic schizophrenia patients: an add-on, double-blind, placebo-controlled study. *European Neuropsychopharmacology*, **13**, 123–8.
37. Araneda, R. and Andrade, R. (1991). 5-Hydroxytryptamine2 and 5-hydroxitryptamine1A receptors mediate opposing responses on membrane excitability in rat association cortex. *Neuroscience*, **40**, 399–412.
38. Ichikawa, J., Meltzer, H.Y. (1999). R(+)-8-OH-DPAT, a serotonin1A receptor agonist, potentiated S(-) sulpiride-induced dopamine release in rat medial prefrontal cortex and nucleus accumbens but not striatum. *Journal of Pharmacology and Experimental Therapeutics*, **291**, 1227–32.
39. Sakaue, M., Somboonthum, P., Nishihara, B., *et al.* (2000). Postsynaptic 5-hydroxytryptamine(1A) receptor activation increases in vivo dopamine release in rat prefrontal cortex. *British Journal of Pharmacology*, **129**, 1028–34.
40. Sumiyoshi, T., Matsui, M., Yamashita, I., *et al.* (2001). Enhancement of cognitive performance in schizophrenia by addition of tandospirone to neuroleptic treatment. *American Journal of Psychiatry*, **158**, 1722–5.
41. Sumiyoshi, T., Jayathilake, K., Roy, A., *et al.* (2006). Effect of buspirone, a serotonin(1A) partial agonist, on cognitive function in schizophrenia. [abstract]. *International Journal of Neuropsychopharmacology*, **9** (suppl 1), S248.

42. Bonaccorso, S., Meltzer, H.Y., Li, Z., *et al.* (2002). SR46349-B, a 5-HT(2A/2C) receptor antagonist, potentiates haloperidol-induced dopamine release in rat medial prefrontal cortex and nucleus accumbens. *Neuropsychopharmacology*, **27**, 430–41.
43. Kroeze, W.K., Hufeisen, S.J., Popadak, B.A., *et al.* (2003). H1-histamine receptor affinity predicts short-term weight gain for typical and atypical antipsychotic drugs. *Neuropsychopharmacology*, **28**, 519–26.
44. Reynolds, G.P., Templeman, L.A., Zhang, Z.J. (2005). The role of 5-HT2C receptor polymorphisms in the pharmacogenetics of antipsychotic drug treatment. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **29**, 1021–8.
45. Bigliani, V., Mulligan, R.S., Acton, P.D., *et al.* (2000). Striatal and temporal cortical D2/D3 receptor occupancy by olanzapine and sertindole in vivo: a [123I] epidepride single photon emission tomography (SPET) study. *Psychopharmacologia*, **150**, 132–40.
46. Stephenson, C.M., Bigliani, V., Jones, H.M., *et al.* (2000). Striatal and extra-striatal D(2)/D(3) dopamine receptor occupancy by quetiapine in vivo. *British Journal of Psychiatry*, **177**, 408–15.
47. Xiberas, X., Martinot, J.L., Mallet, L., *et al.* (2001). Extrastriatal and striatal D(2) dopamine receptor blockade with haloperidol or new antipsychotic drugs in patients with schizophrenia. *British Journal of Psychiatry*, **179**, 503–8.
48. Bressan, R.A., Erlandsson, K., Stone, J.M., *et al.* (2005). Impact of schizophrenia and chronic antipsychotic treatment on [123I] CNS-1261 binding to N-methyl-Daspartate receptors in vivo. *Biological Psychiatry*, **58**, 41–6.
49. Kane, J. and Marder, S. (1993). Psychopharmacologic treatment of schizophrenia. *Schizophrenia Bulletin*, **19**, 287–302.
50. Lieberman, J., Jody, D., Geisler, S., *et al.* (1993). Time course and biologic correlates of treatment response in fi rst–episode schizophrenia. *Archives of General Psychiatry*, **50**, 369–76.
51. Meltzer, H., Lee, M., and Colal, P. (1998). The evolution of treatment resistance. Biological implications. *Journal of Clinical Psychopharmacology*, **18**, 5–11.
52. Meltzer, H. (1997). Treatment-resistant schizophrenia: the role of clozapine. *Current Medical Research Opinion*, **14**, 1–20.
53. Tandon, R., Ribeiro, S., DeQuardo, J., *et al.* (1993). Covariance of positive and negative symptoms during neuroleptic treatment in schizophrenia: a replication. *Biological Psychiatry*, **34**, 495–7.
54. The Collaborative Working Group (1998). Assessing the effects of a typical antipsychotics on negative symptoms. *Journal of Clinical Psychiatry*, **12**, 28–35.
55. Saykin, A., Shtasel, D., Gur, R., *et al.* (1994). Neuropsychological deficits in neuroleptic naïve patients with first–episode schizophrenia. *Archives of General Psychiatry*, **51**, 124–31.
56. Palmer, B., Heaton, R., Paulsen, J., *et al.* (1997). Is it possible to be schizophrenic yet neuropsychologically normal? *Neuropsychology*, **11**, 437–46.
57. Green, M. (1996). What are the functional consequences of neurocognitive deficits in schizophrenia? *American Journal of Psychiatry*, **153**, 321–30.
58. Meltzer, H.Y., and McGurk, S. (1999). The effect of clozapine, risperidone and olanzapine on cognitive function in schizophrenia. *Schizophrenia Bulletin*, **25**, 233–56.
59. Sheitman, B., Lee, H., Strauss, R., *et al.* (1997). The evaluation and treatment of first–episode psychosis. *Schizophrenia Bulletin*, **23**, 653–61.
60. McEvoy, J., Hogarty, G., and Steingard, S. (1991). Optimal dose of neuroleptic in acute schizophrenia. A controlled study of the neuroleptic threshold and higher haloperidol dose. *Archives of General Psychiatry*, **48**, 739–45.
61. Carpenter, W., Buchanan, R., Kirkpatrick, B., *et al.* (1999). Diazepam treatment of early signs of exacerbation in schizophrenia. *American Journal of Psychiatry*, **156**, 299–303.
62. Schooler, N.R. (2003). Relapse and rehospitalization: comparing oral and depot antipsychotics. *Journal of Clinical Psychiatry*, **64** (suppl 16), 14–7.
63. Lehman, A.F., Kreyenbuhl, J., Buchanan, R.W., *et al.* (2004). The Schizophrenia Patient Outcomes Research Team (PORT): updated treatment recommendations 2003. *Schizophrenia Bulletin*, **30**, 193–217.
64. Lehman, A.F., *et al*. American Psychiatric Association; Steering Committee on Practice Guidelines. (2004). Practice guideline for the treatment of patients with schizophrenia, second edition. *American Journal of Psychiatry*, **161** (2 suppl), 1–56.
65. Lieberman, J.A., Stroup, T.S., *et al.* Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators. (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*, **353**, 1209–23.
66. Jones, P.B., Barnes, T.R., Davies, L., *et al*. (2006). Randomized controlled trial of the effect on Quality of Life of second- vs first generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Archives of General Psychiatry*, **63**, 1079–87.
67. Revicki, D. (1999). Pharmacoeconomic studies of atypical antipsychotic drugs for the treatment of schizophrenia. *Schizophrenia Research*, **35**, 101–9.
68. Bobes, J., Canas, F., Rejas, J., *et al.* (2004). Economic consequences of the adverse reactions related with antipsychotics: an economic model comparing tolerability of ziprasidone, olanzapine, risperidone, and haloperidol in Spain. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **28**, 1287–97.
69. Meltzer, H. (1979). The clozapine story. In: *The handbook of psychopharmacology trials* (eds. M. Hertzman and D. Feltner), pp. 137–56. New York University Press.
70. Baldessarini, R. and Frankenberg, F. (1991). Clozapine: a novel antipsychotic agent. *New England Journal of Medicine*, **324**, 746–54.
71. Kane, J., Honigfeld, G., *et al.* Clozaril Collaborative Study Group (1988). Clozapine for the treatmentresistant schizophrenic: a double–blind comparison with chlorpromazine. *Archives of General Psychiatry*, **45**, 789–96.
72. Meltzer, H.Y., Alphs, L., *et al.* International Suicide Prevention Trial Study Group. (2003). Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT). *Archives of General Psychiatry*, **60**, 82–91.
73. Nordstrom, A., Farde, L., Nyber, S., *et al.* (1995). D1, D2, and 5-HT2 receptor occupancy in relation to clozapine serum concentration: a PET study of schizophrenic patients. *American Journal of Psychiatry*, **152**, 1444–9.
74. Shiloh, R., Zemishlany, Z., Aizenberg, D., *et al.* (1997). Sulpiride augmentation in people with schizophrenia partially responsive to clozapine. A double-blind, placebo-controlled study. *British Journal of Psychiatry*, **171**, 569–73.
75. Havaki-Kontaxaki, B.J., Ferentinos, P.P., Kontaxakis, V.P., *et al.* (2006). Concurrent administration of clozapine and electroconvulsive therapy in clozapine-resistant schizophrenia. *Clinical Neuropharmacology*, **29**, 52–6.
76. Meltzer, H., Lee, M., Ranjan, R., *et al.* (1996). Relapse following clozapine withdrawal: effect of cyproheptadine plus neuroleptic. *Psychopharmacology*, **124**, 176–87.
77. Marder, S. and Meibach, R. (1994). Risperidone in the treatment of schizophrenia. *American Journal of Psychiatry*, **151**, 825–35.
78. Pajonk, F.G. (2004). Risperidone in acute and long-term therapy of schizophrenia—a clinical profile. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **28**, 15–23.
79. Leucht, S., Pitschel–Walz, G., Abraham, D., *et al.* (1999). Efficacy and extrapyramidal side–effects of the new antispychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta–analysis of randomized controlled trials. *Schizophrenia Research*, **35**, 51–68.

80. Citrome, L., Bilder, R.M., Volavka, J. (2002). Managing treatmentresistant schizophrenia: evidence from randomized controlled trials. *Journal of Psychiatric Practice*, **8**, 205–15.
81. Kumar, V. and Brecher, M. (1999). Psychopharmacology of atypical antipsychotics and clinical outcomes in elderly patients. *Journal of Clinical Psychiatry*, **60**, 10–16.
82. Simpson, G. and Lindenmayer, J. (1997). Extrapyramidal symptoms in patients treated with risperidone. *Journal of Clinical Psychopharmacology*, **17**, 194–201.
83. Csernansky, J.G., Mahmoud, R., Brenner, R. (2002). A comparison of risperidone and haloperidol for the prevention of relapse in patients with schizophrenia. *New England Journal of Medicine*, **346**, 16–22.
84. Rabinowitz, J., Lichtenberg, P., Kaplan, Z., *et al.* (2001). Rehospitalization rates of chronically ill schizohprenic patients discharged on a regimen of risperidone, olanzapine, or conventional antipsychotics. *American Journal of Psychiatry*, **158**, 266–9.
85. Ehret, M.J., Fuller, M.A. (2004). Long-acting injectable risperidone. *Annals of Pharmacotherapy*, **38**, 2122–7.
86. Jeste, D., Lacro, J., Bailey, A., *et al*. (1999). Lower incidence of tardive dyskinesia with risperidone compared with haloperidol in older patients. *Journal of the American Geriatrics Society*, **47**, 716–19.
87. Moller, H., Muller, H., Borison, R., *et al*. (1995). A path-analytical approach to differentiate between direct and indirect drug effects on negative symptoms in schizophrenic patients (a re–evaluation of the North American risperidone study). *European Archives of Psychiatry Clinical Neuroscience*, **245**, 45–9.
88. Carman, J., Peuskens, J., Vangeneugden, A. (1995). Risperidone in the treatment of negative symptoms of schizophrenia: a meta-analysis. *International Clinical Psychopharmacology*, **10**, 207–13.
89. Glennie, J. (1997). Technology overview: pharmaceuticals: pharmacoeconomic evaluations of clozapine in treatment-resistant schizophrenia and risperidone in chronic schizophrenia. Ottawa (ON): Canadian Coordinating Office for Health Technology Assessment (CCOHTA).
90. Laux, G., Heeg, B., van Hout, B.A., *et al*. (2005). Costs and effects of long-acting risperidone compared with oral atypical and conventional depot formulations in Germany. *Pharmacoeconomics*, **23** (suppl 1), 49–61.
91. Awad, A.G. and Voruganti, L.N.P. (2004). Impact of atypical antipsychotics on quality of life in patients with schizophrenia. *CNS Drugs*, **18**, 877–93.
92. Tollefson, G. and Kuntz, A. (1999). Review of recent clinical studies with olanzapine. *British Journal of Psychiatry*, **37**, 30–35.
93. Conley, R., Kelly, D., and Gale, E. (1998). Olanzapine response in treatment-refractory schizophrenic patients with a history of substance abuse. *Schizophrenia Research*, **33**, 95–101.
94. Tollefson, G.D., Birkett, M.A., *et al*., Lilly Resistant Schizophrenia Study Group. (2001). Double-blind comparison of olanzapine versus clozapine in schizohprenic patients clinically eligible for treatment with clozapine. *Biological Psychiatry*, **49**, 52–63.
95. Dinakar, H.S., Sobel, R.N., Bopp, J.H., *et al*. (2002). Efficacy of olanzapine and risperidone for treatment-resistant schizophrenia among long-stay state hospital patients. *Psychiatric Services*, **53**, 755–7.
96. Beasley, C., Tollefson, G., Tran, P., *et al*. (1996). Olanzapine versus placebo and haloperidol acute phase results of the North American double-blind olanzapine trial. *Neuropsychopharmacology*, **14**, 111–23.
97. Tollefson, G. and Sanger, T. (1997). Negative symptoms, a path analytic approach to a double-blind, placebo- and haloperidol-controlled clinical trial with olanzapine. *American Journal of Psychiatry*, **54**, 466–74.
98. Tran, P., Dellva, M., Tollefson, G., *et al.* (1998). Oral olanzapine versus oral haloperidol in the maintenance treatment of schizophrenia and related psychoses. *British Journal of Psychiatry*, **172**, 499–505.
99. Tollefson, G., Sanger, T., Lu, Y., *et al*. (1998). Depressive signs and symptoms in schizophrenia: a prospective blinded trial of olanzapine and haloperidol. *Archives of General Psychiatry*, **55**, 250–8.
100. Almond, S., O'Donnell, O. (1998). Cost analysis of the treatment of schizophrenia in the UK: a comparison of olanzapine and haloperidol. *Pharmacogenomics*, **13**, 575–88.
101. Hamilton, S.H., Revicki, D.A., Edgell, E.T., *et al.* (1999). Clinical and economic outcomes of olanzapine compared with haloperidol for schizophrenia: results from a randomized clinical trial. *Pharmacoeconomics*, **15**, 469–80.
102. Kinon, B.J., Ahl, J., Stauffer, V.L., *et al*. (2004). Dose response and atypical antipsychotics in schizophrenia. *CNS Drugs*, **18**, 597–616.
103. Wagstaff, A.J., Easton, J. and Scott, L.J. (2005). Intramuscular olanzapine: a review of its use in the management of acute agitation. *CNS Drugs*, **19**, 147–64.
104. Arvanitis, L. and Miller, B. (1997). Multiple fixed dose of 'Seroquel' (quetiapine) in patients with acute exacerbation of schizophrenia: a comparison with haloperidol and placebo. The Seroquel Trial 13 Study Group. *Biological Psychiatry*, **42**, 233–46.
105. Small, J., Hirsch, S., *et al*. and the Seroquel Study Group (1997). Quetiapine in patients with schizophrenia: a high- and low-dose double-blind comparison with placebo. *Archives of General Psychiatry*, **54**, 549–57.
106. Gefvert, O., Lundberg, T., Wieselgren, I.M., *et al.* (2001). D(2) and 5HT(2A) receptor occupancy of different doses of quetiapine in schizophrenia: a PET study. *European Neuropsychopharmacology*, **11**, 105–10.
107. Emsley, R.A., Buckley, P., Jones, A.M., *et al.* (2003). Differential effect of quetiapine on depressive symptoms in patients with partially responsive schizophrenia. *Journal of Psychopharmacology*, **17**, 210–15.
108. Velligan, D.I., Newcomer, J., Pultz, J., *et al.* (2002). Does cognitive function improve with quetiapine in comparison to haloperidol? *Schizophrenia Research*, **53**, 239–48.
109. Fernandez, H.H., Trieschmann, M.E., Friedman, J.H. (2003). Treatment of psychosis in Parkinson's disease: safety considerations. *Drug Safety*, **26**, 643–59.
110. Stahl, S.M. and Shayegan, D.K. (2003). The psychopharmacology of ziprasidone: receptor-binding properties and real-world psychiatric practice. *Journal of Clinical Psychiatry*, **64** (Suppl 19), 6–12.
111. Davis, R. and Markham, A. (1997). Ziprasidone. *CNS Drugs*, **8**, 153–9.
112. Tandon, R., Harrigan, E., and Zorn, S. (1997). Ziprasidone: a novel antipsychotic with unique pharmacology and therapeutic potential. *Journal of Serotonin Research*, **4**, 159–77.
113. Swainston, H.T. and Scott, L.J. (2006). Ziprasidone: a review of its use in schizophrenia and schizoaffective disorder. *CNS Drugs*, **20**, 1027–52.
114. Arato, M., O'Connor, R., and Meltzer, H. The Ziprasidone Extended Use in Schizophrenia (Zeus) study: a prospective, double-blind, placebo-controlled, 1-year clinical trial. Submitted for publication.
115. Hirsch, S.R., Kissling, W., Bauml, J., *et al*. (2002). A 28-week comparison of ziprasidone and haloperidol in outpatients with stable schizophrenia. *Journal of Clinical Psychiatry*, **63**, 516–23.
116. Daniel, D.G., Zimbroff, D.L., Potkin, S.G., *et al*. (1999). Ziprasidone 80 mg/day and 160 mg/day in the acute exacerbation of schizophrenia and schizoaffective disorder: a 6-week placebo-controlled trial. Ziprasidone Study Group. *Neuropsychopharmacology*, **20**, 491–505.
117. Harvey, P.D. (2003). Ziprasidone and cognition: the evolving story. *Journal of Clinical Psychiatry*, **64** (suppl 19), 33–9.
118. Phillips, G.A., Van Brunt, D.L., Roychowdhury, S.M., *et al.* (2006). The relationship between quality of life and clinical efficacy from a randomized trial comparing olanzapine and ziprasidone. *Journal of Clinical Psychiatry*, **67**, 1397–403.
119. Bernardo, M., Ramon Azanza, J., Rubio-Terres, C., *et al.* (2006). Cost-effectiveness analysis of schizophrenia relapse prevention: an economic evaluation of the ZEUS (Ziprasidone-Extended-Use-In-Schizophrenia) study in Spain. *Clinical Drug Investigation*, **26**, 447–57.

120. Mamo, D., Kapur, S., Shammi, C.M., *et al.* (2004). A PET study of dopamine D2 and serotonin 5-HT1 receptor occupancy in patients with schizophrenia treated with therapeutic doses of ziprasidone. *American Journal of Psychiatry*, **161**, 818–25.
121. Hamelin, B.A., Allard, S., Laplante, L., *et al.* The effect of timing of a standard meal on the pharmacokinetics and pharmacodynamics of the novel atypical antipsychotic agent ziprasidone. *Pharmacotherapy*, **18**, 9–15.
122. Daniel, D.G., Potkin, S.G., Reeves, K.R., *et al.* (2001). Intramuscular (IM) ziprasidone 20 mg is effective in reducing acute agitation associated with psychosis: a double-bline, randomized trial. *Psychopharmacology*, **155**, 128–34.
123. Newcomer, J.W. and Haupt, D.W. (2006). The metabolic effects of antipsychotic medications. *Canadian Journal of Psychiatry*, **51**, 480–91.
124. Haddad, P.M. and Anderson, I.M. (2002). Antipsychotic-related QTc prolongation, torsade de pointes and sudden death. *Drugs*, **62**, 1649–71.
125. Heinrich, T.W., Biblo, L.A. and Schneider, J. (2006). Torsades de pointes associated with ziprasidone. *Psychosomatics*, **47**, 264–8.
126. Harrigan, E.P., Miceli, J.J., Anziano, R., *et al.* (2004). A randomized evaluation of the effects of six antipsychotic agents on QTc, in the absence and presence of metabolic inhibition. *Journal of Clinical Psychopharmacology*, **24**, 62–9.
127. Burris, K.D., Molski, T.F., Xu, C., *et al.* (2002). Aripiprazole, a novel antipsychotic, is a high-affinity partial agonist at human dopamine D2 receptors. *Journal of Pharmacology and Experimental Therapeutics*, **302**, 381–9.
128. Jordan, S., Koprivica, V., Chen, R., *et al.* (2002). The antipsychotic aripiprazole is a potent, partial agonist at the human 5-HT1A receptor. *European Journal of Pharmacology*, **441**, 137–40.
129. El-Sayeh, H.G., Morganti, C. and Adams, C.E. (2006). Aripiprazole for schizophrenia. Systematic review. *British Journal of Psychiatry*, **189**, 102–8.
130. Potkin, S.G., Saha, A.R., Kujawa, M.J., *et al.* (2003). Aripiprazole, an antipsychotic with a novel mechanism of action, and risperidone vs placebo in patients with schizophrenia and schizoaffective disorder. *Archives of General Psychiatry*, **60**, 681–90.
131. Piggott, T.A., Carson, W.H., *et al.*, Aripiprazole Study Group. (2003). Aripiprazole for the prevention of relapse in stabilized patients with chronic schizophrenia: a placebo-controlled 26-week study. *Journal of Clinical Psychiatry*, **64**, 1048–56.
132. Kasper, S., Lerman, M.N., McQuade, R.D., *et al.* (2003). Efficacy and safety of aripiprazole vs. haloperidol for long-term maintenance treatment following acute relapse of schizophrenia. *International Journal of Neuropsychopharmacology*, **6**, 325–37.
133. Kane, J.M., Meltzer, H.Y., *et al.* Aripiprazole Study Group. (2007). Aripiprazole for treatment-resistant schizophrenia: results of a multicenter, randomized, double-blind, comparison study versus perphenazine. *Journal of Clinical Psychiatry*, **68**, 213–23.
134. Tandon, R., Marcus, R.N., Stock, E.G., *et al.* (2006). A prospective, multicenter, randomized, parallel-group, openlabel study of aripiprazole in the management of patients with schizophrenia or schizoaffective disorder in general psychiatric practice: Broad Effectiveness Trial With Aripiprazole (BETA). *Schizophrenia Research*, **84**, 77–89.
135. Chrzanowski, W.K., Marcus, R.N., Torbeyns, A., *et al.* (2006). Effectiveness of long-term aripiprazole therapy in patients with acutely relapsing or chronic, stable schizophrenia: a 52-week, open-label comparison with olanzapine. *Psychopharmacology*, **189**, 259–66.
136. Kern, R.S., Green, M.F., Cornblatt, B.A., *et al.* (2006). The neurocognitive effects of aripiprazole: an open-label comparison with olanzapine. *Psychopharmacology*, **187**, 312–20.
137. Kerwin, R., Millet, B., Herman, E., *et al.* (2007). A multicentre, randomized, naturalistic, open-label study between aripiprazole and standard of care in the management of community treated schizophrenic patients Schizohprenia Trial of Aripiprazole: (STAR) study. *European Psychiatry*, in press.
138. Tran-Johnson, T.K., Stack, D.A., Marcus, R.N., *et al.* (2007). Efficacy and safety of intramuscular aripiprazole in patients with acute agitation: a randomized, double-blind, placebo-controlled trial. *Journal of Clinical Psychiatry*, **68**, 111–9.
139. Kane, J., Canas, F., Kramer, M., *et al.* (2007). Treatment of schizophrenia with paliperidone extended-release tablets: a 6-week placebo-controlled trial. *Schizophrenia Research*, **90**, 147–61.
140. Davidson, M., Emsley, R., Kramer, M., *et al.* (2007). Efficacy, safety and early response of paliperidone extended-release tablets (paliperidone ER): Results of a 6-week, randomized, placebo-controlled study. *Schizophrenia Research*, **93**, 117–30.
141. Kramer, M., Simpson, G., Maciulis, V., *et al.* (2007). Paliperidone extended-release tablets for prevention of symptom recurrence in patients with schizophrenia: a randomized, double-blind, placebo-controlled study. *Journal of Clinical Psychopharmacology*, **27**, 6–14.
142. Meltzer, H.Y., Bobo, W.V., Nuamah, I.F., *et al.* (2008). Efficacy and tolerability of oral paliperidone extended-release tablets in the treatment of acute schizophrenia: pooled data from three 6-week, placebo-controlled studies. *Journal of Clinical Psychiatry*, **69**, 817–29.
143. McKeage, K., Plosker, G.L. (2004). Amisulpride: a review of its use in the management of schizophrenia. *CNS Drugs*, **18**, 933–56.
144. Möller, J., Boyer, P., *et al.* and the PROD–ASLP Study Group. (1997). Improvement of acute exacerbations of schizophrenia with amisulpride: a comparison with haloperidol. *Psychopharmacology*, **132**, 396–401.
145. Freeman, H. (1997). Amisulpride compared with standard neuroleptics in acute exacerbations of schizophrenia: three efficacy studies. *International Clinical Psychopharmacology*, **12**, 11–17.
146. Paillere–Martinot, M., Lecrubier, Y., Martinot, J., *et al.* (1995). Improvement of some schizophrenic defi cit symptoms with low doses of amisulpride. *American Journal of Psychiatry*, **152**, 130–4.
147. Boyer, P., Lecrubier, Y., Pucch, A., *et al.* (1995). Treatment of negative symptoms in schizophrenia with amisulpride. *British Journal of Psychiatry*, **166**, 68–72.
148. Perry, P., Miller, D., Arndt, S., *et al.* (1992). Clozapine and norclozapine plasma concentrations and clinical response of treatment–refractory schizophrenic patients. *American Journal of Psychiatry*, **148**, 231–5.
149. Peuskens, J., Moller H.J. and Puech, A. (2002). Amisulpride improves depressive symptoms in acute exacerbations of schizophrenia: comparison with haloperidol and risperidone. *European Neuropsychopharmacology*, **12**, 305–10.
150. Colonna, L., Saleem, P., Dondey-Nouvel, L., *et al.* (2000). Long-term safety and effi cacy of amisulpride in subchronic or chronic schizophrenia. Amisulpride Study Group. *International Clinical Psychopharmacology*, **15**, 13–22.
151. Surguladze, S., Patel, A., Kerwin, R.W., *et al.* (2005). Cost analysis of treating schizophrenia with amisulpride: naturalistic mirror image study. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **29**, 517–22.
152. Kelleher, J.P., Centorrino, F., Albert, M.J., *et al.* (2002). Advances in atypical antipsychotics for the treatment of schizophrenia: new formulations and new agents. *CNS Drugs*, **16**, 249–61.
153. Bell, R., McLaren, A., Glanos, J., *et al.* (1998). The clinical use of plasma clozapine levels. *Australian and New Zealand Journal of Psychiatry*, **32**, 567–74.
154. Prior, T.I., Chue, P.S., Tibbo, P., *et al.* (1999). Drug metabolism and atypical antipsychotics. *European Neuropsychopharmacology*, **9**, 301–9.
155. Olesen, O.V., Linnet, K. (2001). Contributions of five human cytochrome P450 isoforms to the N-demethylation of clozapine in vitro at low and high concentrations. *Journal of Clinical Pharmacology*, **41**, 823–32.
156. He, H. and Richardson, J. (1995). A pharmacological, pharmacokinetic and clinical overview of risperidone, a new antipsychotic that blocks serotonin 5–HT2 and dopamine D2 receptors. *International Clinical Psychopharmacology*, **10**, 19–30.

157. Aravagiri, M., Ames, D., Wirshing, W., *et al.* (1997). Plasma level monitoring of olanzapine in patients with schizophrenia: determination by high–performance liquid chromatography with electrochemical detection. *Therapeutic Drug Monitoring*, **19**, 307–13.
158. Casey, D. (1996). 'Seroquel'(quetiapine): preclinical and clinical findings of a new atypical antipsychotic. *Experimentation, Opinion and Investigation of Drugs*, **5**, 939–57.
159. Potkin, S.G., Thyrum, P.T., Alva, G., *et al.* (2002). The safety and pharmacokinetics of quetiapine when coadministered with haloperidol, risperidone, or thioridazine. *Journal of Clinical Psychopharmacology*, **22**, 121–30.
160. Chengappa, K.N., Parepally, H., Brar, J.S., *et al.* (2003). A random-assigment, double-blind, clinical trial of once- vs twice-daily administration of quetiapine fumarate in patients with schizophrenia or schizoaffective disorder: a pilot study. *Canadian Journal of Psychiatry*, **48**, 198–94.
161. Gefvert, O., Bergstrom, M., Langstrom, B., *et al.* (1998). Time course of central nervous dopamine-D2 and 5-HT2 receptor blockade and plasma drug concentrations after discontinuation of quetiapine (Seroquel) in patients with schizophrenia. *Psychopharmacology*, **135**, 119–26.
162. Beedham, C., Miceli, J.J., Obach, R.S. (2003). Ziprasidone metabolism, aldehyde oxidase, and clinical implications. *Journal of Clinical Psychopharmacology*, **23**, 229–32.
163. Swainston, H.T. and Perry, C.M. (2004). Aripiprazole: a review of its use in schizophrenia and schizoaffective disorder. *Drugs*, **64**, 1715–36.
164. Jung, S.M., Kim, K.A., Cho, H.K., *et al.* (2005). Cytochrome P450 3A inhibitor itraconazole affects plasma concentrations of risperidone and 9-hydroxyrisperidone in schizophrenic patients. *Clinical Pharmacology and Therapeutics*, **78**, 520–8.
165. Sachdev, P. (1995). The epidemiology of drug-induced akathisia: Part I. Acute akathisia. *Schizophrenia Bulletin*, **21**, 431–49.
166. Atbasoglu, E.C., Schultz, S.K. and Andreasen, N.C. (2001). The relationship of akathisia with suicidality and depresonalization among patients with schizophrenia. *Journal of Neuropsychiatry and Clinical Neuroscience*, **13**, 336–41.
167. Cavallaro, R. and Smeraldi, E. (1995). Antipsychotic–induced tardive dyskinesia. *CNS Drugs*, **4**, 278–93.
168. Kane, J.M. (2004). Tardive dyskinesia rates with atypical antipsychotics in adults: prevalence and incidence. *Journal of Clinical Psychiatry*, **65** (suppl 9), 16–20.
169. Velamoor, V. (1998). Neuroleptic malignant syndrome. Recognition, prevention and management. *Drug Safety*, **19**, 73–81.
170. Alvir, J., Lieberman, J., Safferman, A., *et al.* (1993). Clozapine–induced agranulocytosis: incidence and risk factors in the United States. *New England Journal of Medicine*, **329**, 162–7.
171. Lieberman, J. and Safferman, A. (1992). Clinical profile of clozapine: adverse reactions and agranulocytosis. *Psychiatric Quarterly*, **63**, 51–70.
172. Merrill, D.B., Dec, G.W. and Goff, D.C. (2005). Adverse cardiac effects associated with clozapine. *Journal of Clinical Psychopharmacology*, **25**, 32–41.
173. Collaborative Working Group. (1998). Adverse effects of the atypical antipsychotics. *Journal of Clinical Psychiatry*, **12**, 17–22.
174. Kelly, D.L., Love, R.C. (2001). Ziprasidone and the QTc interval: pharmacokinetic and pharmacodynamic considerations. *Psychopharmacology Bulletin*, 35, 66–79.
175. Garver, D.L., Holcomb, J.A. and Christensen, J.D. (2005). Cerebral cortical gray expansion associated with two second-generation antipsychotics. *Biological Psychiatry*, **58**, 62–6.
176. Marder, S.R., Kramer, M., Ford, L., *et al.* (2007). Efficacy and safety of paliperidone extended-release tablets: results of a 6-week, randomized, placebo-controlled study. Biological Psychiatry, **62**, 1363–70.
177. Papakostas, G.I., Shelton, R.C., Smith, J., *et al.* (2007). Augmentation of antidepressants with atypical antipsychotic medication for treatment-resistant major depressive disorder: a meta-analysis. *Journal of Clinical Psychiatry*, **68**, 826–31.

6.2.6 Antiepileptic drugs

Brian P. Brennan and Harrison G. Pope Jr.

Introduction

Several drugs originally developed to treat epilepsy have been found effective in certain psychiatric disorders. This chapter reviews the antiepileptic drugs most extensively studied in psychiatric disorders: valproate, carbamazepine, lamotrigine, and topiramate. We then briefly mention six other antiepileptics currently under investigation in various psychiatric disorders, but not yet extensively studied: gabapentin, oxcarbazepine, levetiracetam, tiagabine, zonisamide, and pregabalin. The antiepileptic drug phenytoin is rarely used in psychiatric disorders, and is therefore not included in this chapter. The benzodiazepines, which have antiepileptic properties, are also omitted here, as they are discussed in Chapter 6.2.2. We briefly list studies documenting the efficacy of these various agents in psychiatric disorders, but the reader is referred to the individual chapters on specific disorders for a more detailed discussion of treatment strategies.

Valproate[1]

Introduction

Valproate (valproic acid) is a simple branched-chain carboxylic acid, first used as an organic solvent in the late 1800s (see Fig. 6.2.6.1). Its antiepileptic properties were discovered serendipitously in 1963, and its clinical use as an antiepileptic drug began in 1964. As early as 1966, valpromide (the amide precursor of valproate) was reported to be effective in the treatment of bipolar disorder.(1) Since then, valproate has been used effectively in the treatment of numerous psychiatric and neurologic conditions, and is now widely used as a mood stabilizer in the treatment of bipolar disorder.

Valproate is currently available as five different preparations: valproate (Depakene), sodium valproate (Depakene syrup), divalproex sodium (Depakote) (which is an equal proportion of sodium valproate and valproic acid), divalproex sodium sprinkle capsules (Depakote sprinkle capsules), and valpromide (the amide precursor of valproate, which is available in Europe, but not in the United States).

Pharmacology

The mechanism of action of valproate in the treatment of epilepsy is unclear, but appears to be related to increased levels of gamma-aminobutyric acid (GABA) in the brain. It inhibits the breakdown and turnover of GABA, increases its release, and increases the density of the GABA-ββ receptor subtype.(2) Its mechanism of action in treating psychiatric disorders is unknown.

[1] Valproate is marketed in the British Commonwealth as 'valproic acid' and as 'sodium valproate' in the US, but these are effectively interchangeable as they both yield valproate in the bloodstream.

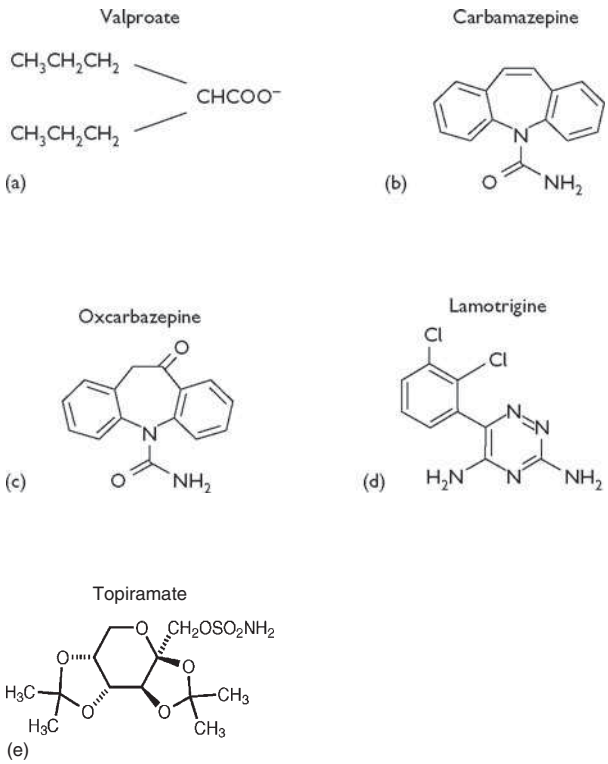


Fig. 6.2.6.1 Molecular structures of selected antiepileptic drugs. (a) Valproate, (b) Carbamazepine, (c) Oxcarbazepine, (d) Lamotrigine, and (e) Topiramate.

Pharmacokinetics

All preparations of valproate are completely absorbed after oral administration. The rate of absorption varies slightly with the different preparations, but these differences are probably not clinically significant. Co-administration with food can delay absorption. Valproate is approximately 90 per cent protein-bound. Only the unbound drug crosses the blood brain barrier and is pharmacologically active in the CNS. As total serum valproate concentration increases, the unbound portion of valproate is disproportionately increased, presumably due to saturation of the protein-binding sites. Therefore, at higher serum concentrations, small increases in dose may result in significant changes in efficacy and side effects. Valproate is metabolized by the liver to a glucuronide conjugate or one of several metabolites, some having antiepileptic activity. The half-life of valproate ranges from 6 to 17 h. Enzyme-inducing antiepileptic drugs, such as carbamazepine and phenytoin, shorten the half-life of valproate (see Interactions).[3]

Side effects

Valproate is often associated with minor side effects, but can rarely cause life-threatening, idiosyncratic reactions. Common side effects include gastrointestinal symptoms, such as nausea, vomiting, abdominal pain, and diarrhoea; and neurological symptoms, such as tremor, somnolence, and dizziness. Weight gain is also common. Hair loss occurs in some patients, but is often transient and reversible upon discontinuation of the drug. Rare, but potentially fatal, idiosyncratic reactions include hepatic failure, acute haemorrhagic pancreatitis, and agranulocytosis.[2,4] Known risk factors for irreversible hepatic failure include young age (especially less than 2 years old), developmental delay, a metabolic disorder, or concomitant administration of other antiepileptic drugs.[5] Because of this risk, liver function tests are recommended prior to initiation of therapy and periodically thereafter (see Dosage and Administration).

Toxic effects

(a) Overdose

Overdose with valproate can result in heart block, coma, and death. Haemodialysis may be useful in clearing the drug rapidly, and naloxone may reverse the CNS depressant effects.[3]

(b) Pregnancy

Valproate increases the risk of neural tube defects (such as spina bifida) to approximately 1–2 per cent of pregnancies when administered in the first trimester. Other reported congenital anomalies include craniofacial defects and cardiovascular malformations. Valproate is found in human breast milk, at approximately 1–10 per cent of serum concentrations, but its effects on the nursing child are unknown.[6]

Indications and contraindications

Controlled trials confirm that valproate is effective in the treatment of multiple seizure types, including complex partial, simple and complex absence, generalized tonic–clonic, and myoclonic seizures.[3] Several controlled studies indicate efficacy in the treatment of acute mania,[7–13] mixed episodes,[12,14] and in the prophylaxis of recurrent mood episodes.[13,15,16] One small controlled study has offered limited evidence for the efficacy of valproate in the treatment of bipolar depression.[17] There is growing support from controlled studies for the efficacy of valproate, combined with both typical and atypical antipsychotics, in the treatment of acute exacerbations of schizophrenia[18,19]; particularly when the presentation includes agitation and hostility.[20] There are also several small controlled studies demonstrating the benefit of valproate in the treatment of mood instability and impulsivity associated with borderline personality disorder.[21–23] Other conditions for which valproate may sometimes be useful include pain syndromes, anxiety disorders, alcohol and sedative withdrawal syndrome, impulse control disorders, and behavioural and affective disturbances associated with intellectual disability and dementia.[2]

Valproate is contraindicated in patients with known hypersensitivity to the drug. It should be used cautiously in patients with significant hepatic disease.

Interactions

In general, valproate can be combined safely with other psychotropic medications and antiepileptic drugs. However, given that valproate is highly protein-bound and can inhibit hepatic enzymes, some drug–drug interactions have been identified.[3] **Aspirin**, which is highly protein-bound, elevates the free fraction of valproate, resulting in increased effects of valproate on the CNS. Valproate can displace **diazepam**, **phenytoin**, **carbamazepine**, and **warfarin** from protein-binding sites, resulting in increased activity of these drugs. Co-administration of valproate with **lamotrigine** significantly increases the half-life of the latter and can increase the risk of lamotrigine-induced rashes. When administered with **carbamazepine**, three potential interactions may occur: (1) valproate can increase the concentration of carbamazepine's metabolite, carbamazepine-10,11-epoxide, by inhibiting its further metabolism; (2) carbamazepine may lower the valproate level; and (3) valproate may increase the carbamazepine level.[24] Therefore, close monitoring of serum

concentrations of both drugs is important when they are combined. **Amitriptyline** and **fluoxetine** may increase serum valproate concentrations, possibly by inhibition of valproate metabolism.

Effects of withdrawal

As with other antiepileptic drugs, valproate should be tapered gradually over several weeks to minimize the risk of rebound seizures.

Dosage and administration

Before initiating treatment with valproate, it is advisable to obtain a baseline complete blood count (CBC), liver function tests (LFTs), and if appropriate, a pregnancy test. CBC and liver function tests should be performed monthly for the first 3 months, and, if no abnormalities are found, every 6 to 12 months thereafter. If hepatic transaminase levels increase to more than three times normal, valproate should be discontinued. If the transaminase levels eventually return to baseline and the patient responded to valproate previously, re-challenge can be considered. If hepatic transaminase levels increase, but are less than three times normal, monitoring should be increased to once every 1–2 weeks until transaminase levels stabilize, and then monthly thereafter.[2]

The initial starting dose of valproate in adults is 250 to 1000 mg per day, given in two or three divided doses (see Table 6.2.6.1 for dosage forms). The dose may be increased every 1–3 days depending on the patient's response and tolerance. The usual therapeutic concentration is between 50–100 μg/ml (drawn 12 h after the last dose) for both psychiatric and neurological disorders. Some clinicians give the entire daily dose of valproate at bedtime. In patients with seizure disorders or acute mania, an oral loading strategy can be used.[25] In this situation, the patient receives 20 mg/kg as a bolus on the first day, resulting in rapid achievement of therapeutic levels. However, psychiatric patients who are not acutely manic usually have difficulty tolerating the oral loading strategy.

Table 6.2.6.1 Available dosage forms of antiepileptic drugs

Drug (proprietary name)	Preparation
Valproate	
Valproate (Depakene)	250 mg capsule
Valproate syrup (Depakene syrup)	250 mg/5 ml
Divalproex sodium (Depakote)	125, 250, and 500 mg tablets
Divalproex sodium extended release (Depakote ER)	250, 500 mg tablets
Divlaproex sodium sprinkle capsules	125 mg capsule
Carbamazepine	
Carbamazepine (Tegretol®)	100, 200 mg tablets; 100 mg chewable tablets; suspension of 100 mg/5 ml
Carbamazepine extended-release tablets	
(Tegretol XR®)	100, 200, and 400 mg tablets
(Carbatrol®)	100, 200, and 300 mg capsules
(Equetro®)	100, 200, and 300 mg capsules
Oxcarbazepine (Trileptal®)	150, 300 and 600 mg tablets; suspension of 300 mg/5 ml
Lamotrigine (Lamictal®)	25, 100, 150, and 200 mg tablets; 2, 5, and 25 mg chewable tablets

Carbamazepine

Introduction

Carbamazepine (Tegretol®) is an iminostilbene derivative with a structure similar to the tricyclic antidepressant imipramine (see Fig. 6.2.6.1). It was initially developed as a potential antidepressant in the 1950s, but was found to have antiepileptic and analgesic properties, and has been marketed for the treatment of seizures and pain syndromes since 1963. For many clinicians, it has been the preferred treatment for partial and generalized tonic–clonic seizures, as well as neuropathic pain. Its clinical use in affective disorders began in the early 1970s; since then, it has become widely used in psychiatry.

Pharmacology

Carbamazepine's mechanism of action in the treatment of seizures and pain syndromes is controversial, but probably results from blockade of voltage-sensitive sodium channels or enhancement of gamma-aminobutyric acid (GABA) activity. Its mechanism of action in psychiatric disorders is unknown, and may be different, given that it affects numerous neurotransmitter systems.[26,27]

Pharmacokinetics

Carbamazepine is absorbed slowly, with peak plasma levels occurring 4–5 h after administration of the tablets. Absorption is faster for the carbamazepine liquid, and slower for carbamazepine extended-release tablets. Oral bioavailability is about 80 per cent; plasma protein binding is approximately 75 per cent. The half-life of carbamazepine is variable, as it induces its own metabolism with chronic administration (autoinduction). Initially, the half-life ranges from 18–65 h, but after autoinduction is complete (usually 3–5 weeks), it is decreased to 5–25 h. Children metabolize carbamazepine more rapidly than adults, and therefore require higher doses to achieve similar levels. Carbamazepine is metabolized in the liver by the cytochrome P_{450} system to a wide variety of metabolites, some with antiepileptic activity. The predominant metabolite, carbamazepine-10, 11-epoxide (CBZ-E), is further metabolized by epoxide hydrolase to an inactive form. Most of carbamazepine's metabolites are excreted as glucuronide conjugates in the urine.[28,29]

Side effects

Carbamazepine is generally well tolerated, with less than 5 per cent of patients discontinuing the medication because of adverse effects. Common side effects seen during initiation of treatment include dizziness, ataxia, sedation, nausea, and diplopia. These are often mild in severity, and frequently resolve with continued treatment.

(a) Haematological side effects

Carbamazepine commonly causes a benign suppression of white blood cell count, but in rare cases may cause severe and potentially fatal blood dyscrasias, including agranulocytosis, pancytopenia, and aplastic anaemia. The incidence of these non-dose-related, idiosyncratic reactions has been estimated to range between 1 in 10 000 to 1 in 300 000.[4]

(b) Hepatic toxicity

Carbamazepine is frequently associated with benign transaminase elevations. Very rarely, a non-dose-related, idiosyncratic reaction causes hepatic failure, which can be fatal.

(c) Cardiovascular effects

Carbamazepine slows intracardiac conduction, and is relatively contraindicated in patients with heart block.

(d) Dermatologic effects

Rashes occur in 5–15 per cent of patients. These are usually benign, but rarely lead to exfoliative dermatitis, Stevens–Johnson syndrome, or toxic epidermal necrolysis. Therefore, it is usually recommended that the drug be discontinued if any rash develops.

(e) Endocrine effects

Carbamazepine can exert antidiuretic effects, which result in hyponatremia in 5–40 per cent of patients.[30] Usually, this effect is clinically insignificant.

Carbamazepine can result in decrease in free T_3 and T_4, but clinical hypothyroidism is extremely rare.

Toxic effects

(a) Overdose

Carbamazepine overdose can be fatal. Common symptoms include nystagmus, tremor, ophthalmoplegia, and myoclonus. Life-threatening effects include atrioventricular block, coma, seizures, and respiratory depression.[31]

(b) Pregnancy

Carbamazepine exposure in the first trimester results in neural tube defects in approximately 1 per cent of infants. Craniofacial abnormalities and developmental delay have been reported as well. Carbamazepine is found in breast milk, but its effects on the nursing infant are unknown.[6,32]

Indications and contraindications

Carbamazepine is indicated for the treatment of simple partial, complex partial, and generalized tonic–clonic seizures. It is ineffective against absence seizures, and may even exacerbate them. Carbamazepine is also indicated in the treatment of trigeminal neuralgia and other neuropathic pain syndromes. Several double-blind, placebo-controlled trials confirm carbamazepine's efficacy in treating both the manic and mixed phase of bipolar disorder.[33,34] There is limited evidence demonstrating efficacy in the treatment of either bipolar or unipolar depression. Uncontrolled reports also suggest that carbamazepine may be useful in the treatment of personality disorders, impulse control disorders, and alcohol/sedative withdrawal syndrome.

Carbamazepine is contraindicated in patients with a history of previous bone marrow depression, hypersensitivity to the drug, or hypersensitivity to any of the tricyclic antidepressants (given its structural similarity to imipramine). Its use with monoamine oxidase inhibitors is not recommended, and carbamazepine should be used with caution in patients with cardiac disease.

Interactions

Given that carbamazepine is extensively metabolized by the liver and induces hepatic enzymes, it produces many significant drug–drug interactions (Table 6.2.6.2).[35–37] Many drug levels are reduced by carbamazepine and can become subtherapeutic. Therefore, it is important to monitor concomitantly administered medications, as dosage adjustments may be necessary.

Table 6.2.6.2 Carbamazepine (CBZ)-drug interactions

CBZ decreases drug levels	Drugs that increase CBZ levels
Alprazolam	Acetazolamide
Clobazam	Cimetidine
Clonazepam	Clarithromycin
Clozapine	Danazol
Dicoumarol	Dextropropoxyphene
Doxycycline	Diltiazem
Ethosuximide	Fluoxetine
Fentanyl	Gemfibrozil
Haloperidol	Isoniazid
Imipramine	Itraconazole
Lamotrigine	Ketaconazole
Mesuximide	Loratadine
Methadone	Macrolide antibiotics
Methylprednisolone	Metronidazole
Oral contraceptives (can result in contraceptive failure)	Nicotinamide
Pancuronium	Nicotinic acid
Paracetamol	Propoxyphene
Phensuximide	Remacemide
Phenytoin (can either increase or decrease)	Rifampicin
Prednisolone	Stiripentol
Primidone	Terfenadine
Remacemide	Valproate
Theophylline	Verapamil
Tiagabine	Viloxazine
Topiramate	
Valproate	
Vecuronium	
Warfarin	
CBZ increases drug levels	**Drugs that decrease CBZ levels**
Clomipramine (possibly)	Cisplatin
Phenytoin (can either increase or decrease)	Doxorubicin
Primidone	Felbamate
	Rifampicin
	Phenobarbital
	Phenytoin
	Primidone
	Theophylline

Effects of withdrawal

As with other antiepileptic drugs, carbamazepine should be gradually tapered over several weeks in order to avoid rebound seizures.

Dosage and administration

Carbamazepine is generally initiated at a starting dose of 100–400 mg, taken either as a single dose or two divided doses (see Table 6.2.6.1 for dosage forms). The dose is gradually increased by 100 or 200 mg every 2 weeks as the patient tolerates. The usual therapeutic serum concentration is 4–12 mg/l (20–50 µmol/l), which is measured before the first morning dose. The half-life of

carbamazepine will decrease with chronic administration due to autoinduction, necessitating frequent monitoring of the serum carbamazepine concentrations and continued dosage adjustment in the first 2 months of therapy.

Laboratory screening

Given the risk of severe blood dyscrasias and hepatic failure, some authorities recommend obtaining a CBC and LFTs at the initiation of treatment. These tests are often repeated every 2 weeks for the first few months of treatment, and then every 3 to 6 months thereafter. However, some authorities argue that testing is unnecessary, since idiosyncratic reactions are rare and may occur too rapidly to be detected by routine laboratory monitoring.

Lamotrigine

Introduction

Lamotrigine (Lamictal®) is a phenyltriazine compound, structurally unrelated to other antiepileptic drugs (see Fig. 6.2.6.1). It was introduced in Ireland in 1993 and in the United Kingdom and the United States in 1994.

Pharmacology

Lamotrigine is thought to act by blocking voltage-sensitive sodium channels, and by inhibiting the release of glutamate. In experimental animal seizure models, it has an antiepileptic profile similar to that of phenytoin and carbamazepine.[38]

Pharmacokinetics

The oral bioavailability of lamotrigine approaches 100 per cent, and absorption is unaffected by food. Peak plasma concentrations are reached 2–3 h after an oral dose. The half-life of lamotrigine is approximately 30 h, but is altered by the presence of other antiepileptic drugs (see Interactions). Plasma protein binding is approximately 55 per cent. Lamotrigine is metabolized by the liver to an inactive glucuronide conjugate, and then excreted in the urine. The clearance of lamotrigine may be reduced in patients with renal impairment and Gilbert's syndrome, and these individuals may benefit from dosage reduction.[39]

Side effects

In general, lamotrigine has few side effects and is better tolerated than other antiepileptic drugs. The most common side effects include dizziness, headache, diplopia, ataxia, blurred vision, nausea, somnolence, and rash. The most concerning side effect is skin rash, which can be life-threatening. Approximately 10 per cent of adults develop a rash while taking lamotrigine, but the majority of these are benign. However, about 1 in 1000 will develop a life-threatening rash, such as Stevens–Johnson syndrome or toxic epidermal necrolysis. The incidence of rash is much higher in paediatric patients, occurring in 1 in 50 to 1 in 100 patients; therefore, lamotrigine should be used with caution in patients less than 16 years of age. Starting at a low dose and slowly increasing it can minimize the risk of rash. Co-administration of valproate can increase the risk of rash. Given the difficulty in predicting who will develop a life-threatening rash, lamotrigine is usually discontinued at the first sign of any rash.

There are a few reports of possible idiosyncratic reactions in patients taking lamotrigine. These include disseminated intravascular coagulation, multiorgan failure, and acute hepatic necrosis.[40] It is unclear, however, if these conditions were actually caused by the drug itself.

Toxic effects

(a) Overdose

The few reported cases of overdose on lamotrigine (at doses up to 4000 mg) were not fatal, but resulted in symptoms such as excessive sedation, dizziness, and headache.

(b) Pregnancy

The effects of lamotrigine on human pregnancy and breast-fed infants are unknown.

Indications and contraindications

Several double-blind, placebo-controlled, add-on trials confirm lamotrigine's efficacy in treating some patients with partial or generalized tonic–clonic seizures.[39] Clinical trials also suggest efficacy against absence, atypical absence, and myoclonic seizures, as well as seizures associated with the Lennox–Gastaut syndrome.[41] Lamotrigine demonstrated efficacy in the treatment of bipolar depression in a large placebo-controlled trial.[42] However, it does not appear to be beneficial in the treatment of acute mania, largely due to the drug's long titration schedule. Several controlled studies have also demonstrated efficacy in the maintenance treatment of bipolar disorder.[39,43–45] and in the rapid-cycling subtype of bipolar disorder.[46] In addition to bipolar depression, lamotrigine has shown some benefit in two small placebo-controlled studies when added to selective serotonin reuptake inhibitors (SSRIs) as an augmentation strategy in the treatment of unipolar depression.[47,48] Preliminary evidence from two small placebo-controlled studies also indicates that lamotrigine may have benefit as an augmentation therapy with conventional and atypical antipsychotics in treatment-resistant schizophrenia.[49,50]

Interactions

Lamotrigine does not appear to affect the kinetics of other antiepileptic drugs or oral contraceptives, but its own kinetics are markedly affected by the concomitant administration of other antiepileptic drugs. Valproate inhibits the metabolism of lamotrigine, resulting in a doubling of the half-life to almost 60 h. The enzyme-inducing antiepileptic drugs, such as carbamazepine and phenytoin, decrease the half-life to approximately 15 h. These interactions necessitate dosage adjustments when starting lamotrigine (see Dosage and Administration).[38]

Effects of withdrawal

As with other antiepileptic drugs, lamotrigine should be tapered gradually over several weeks in order to avoid rebound seizures.

Dosage and administration

Lamotrigine must be started at a low dose and increased with caution to a therapeutic dosage in order to minimize the risk of rash. The patient should be informed of the risk of developing a rash, and instructed to contact the physician immediately if one appears. The starting dose depends upon the concomitant administration of other antiepileptic drugs. See Tables 6.2.6.3–6.2.6.5 for the appropriate lamotrigine titration schedules. Table 6.2.6.1 shows dosage forms.

Table 6.2.6.3 Suggested lamotrigine titration schedule for patients taking carbamazepine, phenytoin, phenobarbital, primidone, or rifampicin and not taking valproate

Week 1	25 mg daily
Week 2	50 mg daily
Week 3	100 mg daily, in divided doses
Week 4	100 mg daily, in divided doses
Week 5	200 mg daily, in divided doses
Week 6	300 mg daily, in divided doses
Week 7	Up to 400 mg daily, in divided doses

Table 6.2.6.4 Suggested lamotrigine monotherapy for patients not taking carbamazepine, phenytoin, phenobarbital, primidone, or rifampicin and not taking valproate

Week 1	12.5 mg daily
Week 2	25 mg daily
Week 3	50 mg daily
Week 4	50 mg daily
Week 5	100 mg daily
Week 6	150 mg daily
Week 7	200 mg daily

Table 6.2.6.5 Suggested lamotrigine titration schedule for patients taking valproate

Week 1	12.5 mg every *other* day
Week 2	25 mg every *other* day
Week 3	25 mg daily
Week 4	25 mg daily
Week 5	50 mg daily
Week 6	75 mg daily
Week 7	100 mg daily

Topiramate

Introduction

Topiramate (Topamax®) is a sulfamate-substituted monosaccharide originally developed as an oral hypoglycemic agent. It was subsequently found to have antiepileptic effects. It is currently approved for epilepsy in over 60 countries and for migraine prophylaxis in more than 20 countries.

Pharmacology

Topiramate is believed to act through several different mechanisms including: (1) inhibition of sodium channel conductance; (2) inhibition of L-type calcium channels; (3) increase in GABA release through an unknown mechanism; (4) decrease in glutamate-mediated excitation through blockade of kainate receptors; and (5) inhibition of carbonic anhydrase.[51]

Pharmacokinetics

There is almost complete, linear, and rapid absorption of topiramate across dose ranges.[52] The absorption of topiramate is not affected by food. It is not significantly protein-bound. Hepatic metabolism (~20 per cent) is less important than renal clearance (~80 per cent) and inactive metabolites comprise less than 5 per cent the administered dose.[52] The normal half-life of topiramate is 19–23 h; this is not influenced by the administration of hepatic enzyme-inducing medications such as carbamazepine or phenytoin.[52]

Side effects

The most common side effects of topiramate are memory and concentration difficulties, paresthesias, somnolence, dizziness, anorexia, and weight loss. In clinical trials with topiramate, kidney stones are reported 2–4 times more frequently than in the general population, probably as a result of the drug's inhibition of carbonic anhydrase. Patients predisposed to developing kidney stones should maintain good hydration during topiramate therapy to minimize the risk of renal stone formation. Rarely, topiramate has been associated with acute myopia precipitating secondary angle closure glaucoma and with oligohidrosis leading to hyperthermia.

Toxic effects

(a) Overdose

Several cases of topiramate overdose have been recorded, with signs and symptoms including severe metabolic acidosis, convulsions, drowsiness, speech disturbance, blurred vision, diplopia, lethargy, tupor, hypotension, abdominal pain, agitation, and dizziness. The clinical consequences in most cases were not severe. However, deaths have occurred after poly-drug overdoses involving topiramate.

(b) Pregnancy

The effects of topiramate on human pregnancy and breast-fed infants are unknown.

Indications and contraindications

Topiramate has confirmed efficacy as monotherapy for partial-onset seizures or primary generalized tonic–clonic seizures for patients 10 years and older and as adjunctive therapy for the same seizure types, including seizures associated with Lennox–Gastault syndrome, in patients of age 2 and older.[52] In addition, several large controlled studies have demonstrated benefit in the prophylaxis of migraines.[53–55] It may also be helpful in the treatment of essential tremor and chronic pain syndromes. There are no large placebo-controlled studies investigating the use of topiramate as monotherapy for bipolar disorder. However, there is limited evidence from several small open-label studies for topiramate as an adjunctive treatment for both the depressive[56] and manic[57,58] phases of bipolar disorder. Controlled studies have demonstrated efficacy in eating disorders including binge eating disorder[52] and bulimia nervosa.[59] Given its ability to reduce weight, the utility of topiramate may be greatest as an adjunctive treatment in bipolar disorder with comorbid obesity. Several small placebo-controlled studies have also suggested that topiramate may have benefit in the treatment of alcohol dependence,[60] impulse control disorders such as pathological gambling,[61] and borderline personality disorder.[62]

Interactions

Topiramate does not typically alter the metabolism of any other drugs. One important exception to this is that the effectiveness of

the ethinylestradiol component of oral contraceptives is reduced when more than 200 mg of topiramate is prescribed daily.[52] Topiramate occasionally leads to a modest increase in phenytoin concentrations (0–25 per cent) and to modest increases in the clearance of risperidone, pioglitazone, and lithium.[52] Hydrochlorothiazide can lead to modest increases in serum concentration of topiramate.[52] Concomitant administration of topiramate and valproate has been associated with hyperammonemia with or without encephalopathy in patients who have previously tolerated either drug alone.

Effects of withdrawal

As with other antiepileptic drugs, topiramate should be tapered gradually over several weeks in order to avoid rebound seizures.

Dosage and administration

The recommended dose for monotherapy treatment of epilepsy in adults and children 10 years of age or older is 400 mg/day in two divided doses. It is recommended that therapy be initiated at 25–50 mg/day followed by titration to an effective dose in increments of 25–50 mg/week. The recommended total daily dose for migraine prophylaxis is 100 mg/day administered in two divided doses. In patients with renal impairment one-half of the usual adult dose is recommended.

Other antiepileptics

Several other drugs with antiepileptic activity are currently under investigation in various psychiatric disorders, but have not as yet shown well-documented efficacy. Gabapentin (Neurontin®), a structural analog of gamma-aminobutyric acid (GABA), has been marketed for use as adjunctive therapy in the treatment of epilepsy since 1993 and also carries a clinical indication for postherpetic neuralgia. It displays a pharmacological profile similar to phenytoin and carbamazepine in animal seizure models.[63]

Several placebo-controlled trials confirm gabapentin's efficacy as adjunctive therapy in some patients with partial seizures, especially complex partial seizures and partial seizures with secondary generalization.[42] Additionally, several controlled studies have demonstrated benefit in the treatment of various pain syndromes, particularly neuropathic pain.[64,65] In a large, multicentre, controlled study gabapentin failed as an adjunctive treatment with lithium and/or valproate in the treatment of acute manic or mixed episodes, and actually performed significantly more poorly than placebo.[66] Since that time, however, a small controlled study has suggested that gabapentin may have efficacy as an adjunctive maintenance treatment for bipolar disorder.[67] Two small controlled studies have also demonstrated efficacy for gabapentin in the treatment of panic disorder[68] and social anxiety[69] respectively.

Oxcarbazepine (Trileptal®), the 10-keto analog of carbamazepine, has an antiepileptic effect similar to carbamazepine, but has fewer side effects, and has not been associated with severe blood dyscrasias. The drug also has fewer effects on hepatic enzyme activity, and hence causes fewer drug interactions. Given its similarity to carbamazepine, oxcarbazepine might be expected to be effective for the manic phase of bipolar disorder. However, despite some promising uncontrolled observations, there are as yet no large controlled trials of oxcarbazepine as monotherapy in adult bipolar disorder. A large double-blind study of oxcarbazepine in children with manic or mixed episodes did not demonstrate efficacy.[70] Preliminary evidence from one small open-label study suggests that oxcarbazepine may also reduce the risk of relapse prevention in alcohol abuse.[71]

Zonisamide (Zonegran®) is an antiepileptic that causes carbonic anhydrase inhibition; it is somewhat similar in its pharmacologic profile to topiramate. Like topiramate, it frequently causes weight loss, and has been shown effective for weight loss in a controlled study of obese individuals with binge eating disorder.[72] It also increases the risk of kidney stones to at least the same degree as topiramate. Small open-label studies have suggested that zonisamide may be effective in the treatment of both the manic and depressed phases of bipolar disorder.[73,74]

Tiagabine (Gabitril®) is believed to exert its anticonvulsant effects by enhancing the effects of GABA. This is thought to occur through blockade of reuptake of GABA into presynaptic neurones. Tiagabine showed efficacy in the treatment of generalized anxiety disorder in a large double-blind study.[75] In addition, several small open-label studies have demonstrated preliminary evidence benefit in the treatment of posttraumatic stress disorder,[76] major depressive disorder with comorbid anxiety,[77] and as an augmentation therapy for generalized anxiety disorder.[78] Tiagabine demonstrated limited efficacy as an add-on treatment for treatment-refractory bipolar disorder in one small clinical case series.[79]

Levetiracetam (Keppra®) is a novel anticonvulsant with an unclear mechanism of action. It has demonstrated benefit in the treatment of social anxiety disorder in one small controlled study, warranting further investigation into this potential use.[80]

Pregabalin (Lyrica®) is a structural derivative of GABA. However, it does not act by altering GABA levels or by binding to the GABA receptor. The exact mechanism of action of pregabalin is unknown, although it is most likely related to its high affinity for the alpha$_2$-delta site on voltage-gated calcium channels. Pregabalin has shown promise as an anxiolytic. Three separate large placebo-controlled studies have demonstrated benefit with pregabalin in the treatment of generalized anxiety disorder.[81–83] Pregabalin also demonstrated efficacy in the treatment of social anxiety disorder in one large controlled study.[84]

It is important to note that soon after the completion of this text, the United States Food and Drug Administration released an alert recommending that patients taking antiepileptic drugs be closely monitored for changes in behavior that could indicate worsening of depression or the emergence of suicidal thoughts or behavior. This warning was based on a meta-analysis of placebo-controlled trials involving eleven antiepileptic medications, including both trials in epilepsy and trials in psychiatric disorders. This meta-analysis revealed that patients receiving antiepileptic drugs had approximately twice the risk of suicidal behavior or ideation (0.43%) compared to patients receiving placebo (0.22%).[85] Further investigation and analyses of this possible association are ongoing.

Further information

McElroy, S.L. and Pope, H.G. Jr. (1988). *Use of anticonvulsants in psychiatry: recent advances*. Oxford Health Care, Clifton, NJ.

Muzina, D.J., El-Sayegh, S., and Calabrese, J.R. (2002). Antiepileptic drugs in psychiatry: focus on randomized controlled trials. *Epilepsy Research*, **50**(1–2), 195–202.

Weisler, R.H., Cutler, A.J., Ballenger, J.C., *et al.* (2006). The use of antiepileptic drugs in bipolar disorders: a review based on evidence from controlled trials. *CNS Spectrums*, **11**(10), 788–99.

Ovsiew, F. (2004). Antiepileptic drugs in psychiatry. *Journal of Neurology, Neurosurgery, and Psychiatry*, **75**(12), 1655–8.

References

1. Lambert, P.A., Carraz, G., Borselli, S., *et al.* (1966). Action neuropsychotrope d'un nouvel anti-épileptique: le dépamide. *Annales Medico-psychologiques*, **124**, 707–10.
2. Pope, H.G. Jr. and McElroy, S.L. (1995). Valproate. In *Comprehensive textbook of psychiatry/VI* (6th edn) (eds. H. Kaplan and B. Sadock), pp. 2112–20. Williams & Wilkins, Baltimore.
3. Davis, R., Peters, D.H., and McTavish, D. (1994). Valproate. A reappraisal of its pharmacological properties and clinical efficacy in epilepsy. *Drugs*, **47**, 332–72.
4. Tohen, M., Castillo, J., Baldessarini, R.J., *et al.* (1995). Blood dyscrasias with carbamazepine and valproate: a pharmacoepidemiological study of 2,228 patients at risk. *The American Journal of Psychiatry*, **152**, 413–18.
5. Bryant, A.E. III and Dreifuss, F.E. (1996). Valproate hepatic fatalities. III. U.S. experience since 1986. *Neurology*, **46**, 465–9.
6. Chang, S.I. and McAuley, J.W. (1998). Pharmacotherapeutic issues for women of childbearing age with epilepsy. *The Annals of Pharmacotherapy*, **32**, 794–804.
7. American Psychiatric Association. (2002). Practice guideline for the treatment of patients with bipolar disorder. *The American Journal of Psychiatry*, **159**(Suppl. 4), 1–50.
8. Pope, H.G. Jr, McElroy, S.L., Keck, P.E. Jr., *et al.* (1991). Valproate in the treatment of acute mania. A placebo-controlled study. *Archives of General Psychiatry*, **48**, 62–8.
9. Kravitz, H. M. and Fawcett, J. (1994). Efficacy of divalproex vs lithium and placebo in mania. *The Journal of the American Medical Association*, **272**, 1005–6.
10. Bowden, C.L., Brugger, A.M., Swann, A.C., *et al.* (1994). Efficacy of divalproex vs lithium and placebo in the treatment of mania. The Depakote mania study group. *The Journal of the American Medical Association*, **271**, 918–24.
11. McElroy, S.L., Keck, P.E., Stanton, S.P., *et al.* (1996). A randomized comparison of divalproex oral loading versus haloperidol in the initial treatment of acute psychotic mania. *The Journal of Clinical Psychiatry*, **57**, 142–6.
12. Freeman, T.W., Clothier, J.L., Pazzaglia, P., *et al.* (1992). A double-blind comparison of valproate and lithium in the treatment of acute mania. *The American Journal of Psychiatry*, **149**, 108–11.
13. Tohen, M., Ketter, T.A., Zarate, C.A., *et al.* (2003). Olanzapine versus divalproex sodium for the treatment of acute mania and maintenance of remission: a 47-week study. *The American Journal of Psychiatry*, **160**(7), 1263–71.
14. Swann, A.C., Bowden, C.L., Morris, D., *et al.* (1997). Depression during mania. Treatment response to lithium or divalproex. *Archives of General Psychiatry*, **54**(1), 37–42.
15. Bowden, C.L., Calabrese, J.R., McElroy, S.L., *et al.* (2000). A randomized, placebo-controlled 12-month trial of divalproex and lithium in treatment of outpatients with bipolar I disorder. Divalproex Maintenance Study Group. *Archives of General Psychiatry*, **57**(5), 481–9.
16. Gyulai, L., Bowden, C.L., McElroy, S.L., *et al.* (2003). Maintenance efficacy of divalproex in the prevention of bipolar depression. *Neuropsychopharmacology*, **28**(7), 1374–82.
17. Davis, L.L., Bartolucci, A., and Petty, F. (2005). Divalproex in the treatment of bipolar depression: a placebo-controlled study. *Journal of Affective Disorders*, **85**(3), 259–66.
18. Casey, D.E., Daniel, D.G., Wassef, A.A., *et al.* (2003). Effect of divalproex combined with olanzapine or risperidone in patients with an acute exacerbation of schizophrenia. *Neuropsychopharmacology*, **28**(1), 182–92.
19. Wassef, A.A., Dott, S.G., Harris, A., *et al.* (2000). Randomized, placebo-controlled pilot study of divalproex sodium in the treatment of acute exacerbations of chronic schizophrenia. *Journal of Clinical Psychopharmacology*, **20**(3), 357–61.
20. Citrome, L., Casey, D.E., Daniel, D.G., *et al.* (2004). Adjunctive divalproex and hostility among patients with schizophrenia receiving olanzapine or risperidone. *Psychiatric Services*, **55**(3), 290–4.
21. Hollander, E., Allen, A., Lopez, R.P., *et al.* (2001). A preliminary double-blind, placebo-controlled trial of divalproex sodium in borderline personality disorder. *The Journal of Clinical Psychiatry*, **62**(3), 199–203.
22. Frankenburg, F.R. and Zanarini, M.C. (2002). Divalproex sodium in the treatment of women with borderline personality disorder and bipolar II disorder: a double-blind placebo-controlled pilot study. *The Journal of Clinical Psychiatry*, **63**(5), 442–6.
23. Hollander, E., Swann, A.C., Coccaro, E.F., *et al.* (2005). Impact of trait impulsivity and state aggression on divalproex versus placebo response in borderline personality disorder. *The American Journal of Psychiatry*, **162**(3), 621–4.
24. Bernus, I., Dickinson, R.G., Hooper, W.D., *et al.* (1997). The mechanism of the carbamazepine-valproate interaction in humans. *British Journal of Clinical Pharmacology*, **44**, 21–7.
25. Keck, P.E. Jr., McElroy, S.L., Tugrul, K.C., *et al.* (1993). Valproate oral loading in the treatment of acute mania. *The Journal of Clinical Psychiatry*, **54**, 305–8.
26. McElroy, S.L. and Keck, P.E. (1995). Antiepileptic drugs. In *The American psychiatric press textbook of psychopharmacology* (eds. A.F. Schwartzberg and C.B. Nemeroff), pp. 351–75. American Psychiatric Press, Inc., Washington, DC.
27. Post, R.M. (1995). Carbamazepine. In *Comprehensive textbook of psychiatry/VI* (6th edn) (eds. H. Kaplan and B. Sadock), pp. 1964–72. Williams & Wilkins, Baltimore.
28. Maxmen, J.S. and Ward, N.G. (1995). *Psychotropic drugs: fast facts* (2nd edn). W.W. Norton and Company, Inc., New York.
29. Levy, R.H. and Kerr, B.M. (1988). Clinical pharmacokinetics of carbamazepine. *The Journal of Clinical Psychiatry*, **49**(Suppl. 4), 58–61.
30. Van Amelsvoort, T., Bakshi, R., Devaux, C., *et al.* (1994). Hyponatremia associated with carbamazepine and oxcarbazepine therapy: a review. *Epilepsia*, **35**, 181–8.
31. Spiller, H.A., Krenzelok, E.P., and Cookson, E. (1990). Carbamazepine overdose: a prospective study of serum levels and toxicity. *Journal of Toxicology. Clinical Toxicology*, **28**, 445–58.
32. Hansen, D.K., Dial, S.L., Terry, K.K., *et al.* (1996). In vitro embryotoxicity of carbamazepine and carbamazepine-10,11-epoxide. *Teratology*, **54**, 45–51.
33. Weisler, R.H., Kalali, A.H., Ketter, T.A., and SPD417 Study Group. (2004). A multicenter, randomized, double-blind, placebo-controlled trial of extended-release carbamazepine capsules as monotherapy for bipolar disorder patients with manic or mixed episodes. *The Journal of Clinical Psychiatry*, **65**(4), 478–84.
34. Weisler, R.H., Keck, P.E. Jr, Swann, A.C., *et al.* (2005). Extended-release carbamazepine capsules as monotherapy for acute mania in bipolar disorder: a multicenter, randomized, double-blind, placebo-controlled trial. *The Journal of Clinical Psychiatry*, **66**(3), 323–30.
35. Ketter, T.A., Post, R.M., and Worthington, K. (1991). Principles of clinically important drug interactions with carbamazepine. I. *Journal of Clinical Psychopharmacology*, **11**, 198–203.
36. Ketter, T.A., Post, R.M., and Worthington, K. (1991). Principles of clinically important drug interactions with carbamazepine. II. *Journal of Clinical Psychopharmacology*, **11**, 306–13.
37. Spina, E., Pisani, F., and Perucca, E. (1996). Clinically significant pharmacokinetic drug interactions with carbamazepine. An update. *Clinical Pharmacokinetics*, **31**, 198–214.

38. Fitton, A. and Goa, K. (1995). Lamotrigine: an update of its pharmacology and therapeutic use in epilepsy. *Drugs*, **50**, 691–713.
39. Goodwin, G.M., Bowden, C.L., Calabrese, J.R., *et al.* (2004). A pooled analysis of 2 placebo-controlled 18-month trials of lamotrigine and lithium maintenance in bipolar I disorder. *The Journal of Clinical Psychiatry*, **65**(3), 432–41.
40. Calabrese, J.R., Bowden, C.L., Sachs, G.S., *et al.* (1999). A double-blind placebo-controlled study of lamotrigine monotherapy in outapatients with bipolar I depression. Lamictal 602 Study Group. *The Journal of Clinical Psychiatry*, **60**(2), 79–88.
41. Bowden, C.L., Calabrese. J.R., Sachs, G., *et al.* (2003). A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently manic or hypomanic patients with bipolar I disorder. *Archives of General Psychiatry*, **60**(4), 392–400.
42. Calabrese, J.R., Bowden, C.L., Sachs, G., *et al.* (2003). A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently depressed patients with bipolar I disorder. *The Journal of Clinical Psychiatry*, **64**(9), 1013–24.
43. McElroy, S.L., Zarate, C.A., Cookson, J., *et al.* (2004). A 52-week, open-label continuation study of lamotrigine in the treatment of bipolar depression. *The Journal of Clinical Psychiatry*, **65**(2), 204–10.
44. Goodwin, G.M., Bowden, C.L., Calabrese, J.R., *et al.* (2004). A pooled analysis of 2 placebo-controlled 18-month trials of lamotrigine and lithium maintenance in bipolar I disorder. *The Journal of Clinical Psychiatry*, **65**(3), 432–41.
45. Chattergoon, D.S., McGuigan, M.A., Koren, G., *et al.* (1997). Multiorgan dysfunction and disseminated intravascular coagulation in children receiving lamotrigine and valproic acid. *Neurology*, **19**, 1442–4.
46. Calabrese, J.R., Suppes, T., Bowden, C.L., *et al.* (2000). A double-blind, placebo-controlled, prophylaxis study of lamotrigine in rapid-cycling bipolar disorder. Lamictal 614 Study Group. *The Journal of Clinical Psychiatry*, **61**(11), 841–50.
47. Normann, C., Hummel, B., Scharer, L.O., *et al.* (2002). Lamotrigine as adjunct to paroxetine in acute depression: a placebo-controlled, double-blind study. *Journal of Clinical Psychiatry*, **63**(4), 337–44.
48. Barbosa, L., Berk, M., and Vorster, M. (2003). A double-blind, randomized, placebo-controlled trial of augmentation with lamotrigine or placebo in patients concomitantly treated with fluoxetine for resistant major depressive episodes. *The Journal of Clinical Psychiatry*, **64**(4), 403–7.
49. Tiihonen, J., Hallikainen, T., Ryynanen, O.P., *et al.* (2003). Lamotrigine in treatment-resistant schizophrenia: a randomized placebo-controlled trial. *Biological Psychiatry*, **54**(11), 1241–8.
50. Kremer, I., Vass, A., Gorelik, I., *et al.* (2004). Placebo-controlled trial of lamotrigine added to conventional and atypical antipsychotics in schizophrenia. *Biological Psychiatry*, **56**(6), 441–6.
51. van Passel, L., Arif, H., and Hirsch, L.J. (2006). Topiramate for the treatment of epilepsy and other nervous system disorders. *Expert Review of Neurotherapeutics*, **6**(1), 19–31.
52. McElroy, S.L., Arnold, L.M., Shapira, N.A., *et al.* (2003). Topiramate in the treatment of binge eating disorder associated with obesity: a randomized, placebo-controlled trial. *The American Journal of Psychiatry*, **160**(2), 255–61.
53. Diener, H.C., Tfelt-Hansen, T., Dahlof, C., *et al.* MIGR-003 study group. (2004). Topiramate in migraine prophylaxis-results from a placebo-controlled trial with propranolol as an active control. *Journal of Neurology*, **251**(8), 943–50.
54. McIntyre, R.S., Mancini, D.A., McCann, S., *et al.* (2002). Topiramate versus bupropion SR when added to mood stabilizer therapy for the depressive phase of bipolar disorder: a preliminary single-blind study. *Bipolar Disorder*, **4**, 207–13.
55. Chengappa, K.N., Rathore, D., Levine, J., *et al.* (1999). Topiramate as add-on treatment for patients with bipolar mania. *Bipolar Disorder*, **1**, 42–53.
56. Bahk, W.M., Shin, Y.C., Woo, J.M., *et al.* (2005). Topiramate and divalproex in combination with risperidone for acute mania: a randomized, open-label study. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, **29**, 115–21.
57. McElroy, S.L., Arnold, L.M., Shapira, N.A., *et al.* (2003). Topiramate in the treatment of binge eating disorder associated with obesity: a randomized, placebo-controlled trial. *The American Journal of Psychiatry*, **160**(2), 255–61.
58. Silberstein, S.D., Schmitt, J., Neto, W., *et al.* (2004). Topiramate in migraine prevention: results of a large controlled trial. *Archives of Neurology*, **61**(4), 490–5.
59. Hoopes, S.P., Reimherr, F.W., Hedges, D.W., *et al.* (2003). Treatment of bulimia nervosa with topiramate in a randomized, double-blind, placebo-controlled trial, 1. Improvement in binge and purge measures. *The Journal of Clinical Psychiatry*, **64**(11), 1335–41.
60. Johnson, B.A., Ait-Daoud, N., Akhtar, F.Z., *et al.* (2004). Oral topiramate reduces the consequences of drinking and improves the quality of life of alcohol-dependent individuals: a randomized controlled trial. *Archives of General Psychiatry*, **61**(9), 905–12.
61. Dannon, P.N., Lowengrub, K., Gonopolski, Y., *et al.* (2005). Topiramate versus fluvoxamine in the treatment of pathological gambling: a randomized, blind-rater comparison study. *Clinical Neuropharmacology*, **28**(1), 6–10.
62. Nickel, M.K., Nickel, C., Mitterlehner, F.O., *et al.* (2004). Topiramate treatment of aggression in female borderline personality disorder patients: a double-blind, placebo -controlled study. *The Journal of Clinical Psychiatry*, **65**(11), 1515–19.
63. Goa, K.L. and Sorkin, E.M. (1993). Gabapentin. *Drugs*, **46**, 409–27.
64. Serpell, M.G. and Neuropathic Pain Study Group. (2003). Gabapentin in neuropathic pain syndromes: a randomized, double-blind, placebo-controlled trial. *Pain*, **99**, 557–66.
65. Morello, C.M., Leckband, S.G., Stoner, C.P., *et al.* (1999). Randomized, double-blind study comparing the efficacy of gabapentin with amitriptyline on diabetic peripheral neuropathy pain. *Archives of Internal Medicine*, **159**, 1931–7.
66. Pande, A.C., Crockatt, J.G., Janney, C.A., *et al.* (2000). Gabapentin in bipolar disorder: a placebo-controlled trial of adjunctive therapy. *Bipolar Disorders*, **2**, 249–55.
67. Vieta, E., Manuel-Goikolea, J., Martinez-Aran, A., *et al.* (2006). A double-blind, randomized, placebo-controlled, prophylaxis study of adjunctive gabapentin for bipolar disorder. *The Journal of Clinical Psychiatry*, **67**, 473–7.
68. Pande, A.C., Pollack, M.H., Crockatt, J., *et al.* (2000). Placebo-controlled study of gabapentin treatment of panic disorder. *Journal of Clinical Psychopharmacology*, **20**(4), 467–71.
69. Pande, A.C., Davidson, J.R., Jefferson, J.W., *et al.* (1999). Treatment of social phobia with gabapentin: a placebo-controlled study. *Journal of Clinical Psychopharmacology*, **19**(4), 341–8.
70. Wagner, K.D., Kowatch, R.A., Emslie, G.J., *et al.* (2006). A double-blind, randomized, placebo-controlled trial of oxcarbazepine in the treatment of bipolar disorder in children and adolescents. *The American Journal of Psychiatry*, **163**(7), 1179–86.
71. Croissant, B., Diehl, A., Klein, O., *et al.* (2006). A pilot study of oxcarbazepine versus acamprosate in alcohol-dependent patients. *Alcoholism, Clinical and Experimental Research*, **30**(4), 630–5.
72. McElroy, S.L., Kotwal, R., Guerdjikova, A.I., *et al.* (2006). Zonisamide in the treatment of binge eating disorder with obesity: a randomized controlled trial. *The Journal of Clinical Psychiatry*, **67**(12), 1897–906.
73. McElroy, S.L., Suppes, T., Keck, P.E. Jr., *et al.* (2005). Open-label adjunctive zonisamide in the treatment of bipolar disorders: a prospective trial. *The Journal of Clinical Psychiatry*, **66**(5), 617–24.
74. Ghaemi, S.N., Zablotsky, B., Filkowski, M.M., *et al.* (2006). An open prospective study of zonisamide in acute bipolar depression. *Journal of Clinical Psychopharmacology*, **26**(4), 385–8.
75. Pollack, M.H., Roy-Byrne, P.P., Van Amerigen, M., *et al.* (2005). The selective GABA reuptake inhibitor tiagibine for the treatment of

generalized anxiety disorder: results of a placebo-controlled study. *The Journal of Clinical Psychiatry*, **66**(11), 1401–8.
76. Connor, K.M., Davidson, J.R., Weisler, R.H., *et al.* (2006). Tiagabine for posttraumatic stress disorder: effects of open-label and double-blind discontinuation treatment. *Psychopharmacology*, **184**(1), 21–5.
77. Carpenter, L.L., Schecter, J.M., Tyrka, A.R., *et al.* (2006). Open-label tiagabine monotherapy for major depressive disorder with anxiety. *The Journal of Clinical Psychiatry*, **67**(1), 66–71.
78. Schwartz, T.L., Azhar, N., Husain, J., *et al.* (2005). An open-label study of tiagabine as augmentation therapy for anxiety. *Annals of Clinical Psychiatry*, **17**(3), 167–72.
79. Suppes, T., Chisholm, K.A., Dhavale, D., *et al.* (2002). Tiagabine in treatment refractory bipolar disorder: a clinical case series. *Bipolar Disorders*, **4**(5), 283–9.
80. Zhang, W., Connor, K.M., and Davidson, J.R. (2005). Levetiracetam in social phobia: a placebo controlled pilot study. *Journal of clinical Psychopharmacology*, **19**(5), 551–3.
81. Feitner, D.E., Crockatt, J.G., Dubovsky, S.J., *et al.* (2003). A randomized, double-blind, placebo-controlled, fixed-dose, multicenter study of pregabalin in patients with generalized anxiety disorder. *Journal of Clinical Psychopharmacology*, **23**(3), 240–9.
82. Rickels, K., Pollack, M.H., Feitner, D.E., *et al.* (2005). Pregabalin for treatment of generalized anxiety disorder: a 4-week, multicenter, double-blind, placebo-controlled trial of pregabalin and alprazolam. *Archives of General Psychiatry*, **62**(9), 1022–30.
83. Montgomery, S.A., Tobias, K., Zornberg, *et al.* (2006). Efficacy and safety of pregabalin in the treatment of generalized anxiety disorder: a 6-week, multicenter, randomized, double-blind, placebo-controlled comparison of pregabalin and venlafaxine. *The Journal of Clinical Psychiatry*, **67**(5), 771–82.
84. Pande, A.C., Feitner, D.E., Jefferson, J.W., *et al.* (2004). Efficacy of the novel anxiolytic pregabalin in social anxiety disorder: a placebo-controlled, multicenter study. *Journal of Clinical Psychopharmacoly*, **24**(2), 1419.
85. United States Food and Drug Administration (January 31, 2008). *Information for Healthcare Professionals - Suicidality and Antiepileptic Drugs.*

6.2.7 Drugs for cognitive disorders

Leslie Iversen

Introduction

Cognitive disorders are among the most difficult of all nervous system illnesses to treat as they affect the most complex and least clearly understood aspects of brain function. Animal studies cannot accurately mirror the complexities of human cognition, and there are few, if any, animal models of human cognitive illnesses. As so few drugs have been found to exert clinically significant effects, animal models for testing novel cognition-enhancing agents have unknown predictive value. However, progress has been made in recent years with improved international agreement on the criteria used to approve new cognition-enhancing drugs, and the introduction of new drugs for the treatment of dementia.

Alzheimer's disease

It is important to define the objective of drug treatment in this, the most common of all forms of senile dementia. Alzheimer's disease (**AD**) is a progressive illness; drug treatment could treat the symptoms without influencing the course of the disease, or it might seek to delay or arrest the progressive cognitive deterioration which such patients suffer. Although the latter aim is the subject of intensive research in academic and industrial laboratories,[1] there are no drugs that target the underlying pathology, and only palliative treatments are, as yet, available.

The approval of new medicines for the symptomatic treatment of AD in recent years has led regulatory agencies to define more clearly what criteria should be used in assessing the clinical benefits derived from drug treatment. AD is a disease characterized by disturbances in higher cortical function, including disorders of recent memory, language function, praxis, visual perception, abstract thinking, and decision making. A variety of composite dementia assessments designed to provide an overall summary of cognitive status, for example the Mini Mental State Examination (**MMSE**), Alzheimer's Disease Assessment Scale-Cognition (**ADAS-Cog**), and the Brief Cognitive Rating Scale are used.[2] Most studies with cholinesterase inhibitors in AD have used ADAS-Cog (a 70-point scale), and a two to three-point improvement for the drug-treated group versus placebo at 6 months has generally been accepted. However, statistically significant, but small, drug-induced improvements in cognitive assessment scores do not necessarily represent a clinically significant improvement to the patient or to their doctor; they must be supplemented by evidence of clinical improvement, using some form of Clinical Global Impression of Change as an outcome measure, usually rated by a clinician on a seven-point scale.

The development of agreed scientific and clinical standards for the approval of new drugs has largely eclipsed most of the older drugs that had been used in the treatment of AD and other dementias, since none of them can meet these standards. The older drugs include a range of cerebral vasodilators (e.g. dihydroergotoxin, papaverine, isoxsuprine, cinnarizine) and the so-called 'nootropics' (e.g. piracetam, oxiracetam, aniracetam), which were widely used in some European countries, as well as the so-called 'metabolic enhancers' (e.g. idebenone and indeloxazine) which were popular for a while in Japan.

Cholinergic agents

Attention has focused instead on the cholinergic agents. The 'cholinergic hypothesis' of dementia. was boosted by the discovery in the 1970s that cholinergic neurones are particularly damaged or absent from the brains of patients dying with AD, and that the extent of damage to the cholinergic system correlates with the severity of dementia in life.[3] In AD the damage appears to be particularly severe in the system of cholinergic neurones located deep in the forebrain in the nucleus basalis of Meynert, whose fibres branch extensively and innervate most areas of the cerebral cortex. This neuronal system forms part of the ascending reticular activating system, which plays a key role in the process of selective attention—essential for the laying down of new memories. Consequently, there has been considerable interest in the possibility that 'cholinergic replacement therapy' might relieve the symptoms of AD, in the same way that dopamine replacement therapy has successfully been employed in the treatment of Parkinson's disease. The most successful approach so far has been the use of inhibitors of the enzyme acetylcholinesterase.

Inhibitors of acetylcholinesterase have been known since the nineteenth century with the discovery of physostigmine, a plant

product used as an arrow-tip poison. Irreversible organophosphate inhibitors of acetylcholinesterase were later developed as chemical warfare agents ('nerve gases'), and for more peaceful uses as insecticides. Despite their colourful past, low doses of this class of compounds have proved effective as cognitive enhancers in a wide range of animal tests, including those in which cholinergic function is deliberately impaired.[4] The first clinical trials in patients with AD were performed with physostigmine, and confirmed that the drug had significant beneficial effects on cognitive performance in AD patients.[5] However, it has limited usefulness because, although it is absorbed rapidly, it has only a very short half-life in plasma. This means that to obtain any sustained cognitive benefit it has to be given in doses that are sufficiently high to elicit a number of adverse side-effects; thus, the therapeutic window was very narrow.

Subsequently four other cholinesterase inhibitors with improved profiles have gained approval for use in AD: **tacrine**, **donepezil, rivastigmine, and galantamine**, but tacrine is no longer actively marketed owing to liver toxicity. Clinical data from several thousand patients with AD involved in trials with these cholinesterase inhibitors are now available.[6–8] The first of these to gain approval in 1997 was donepezil. Results of large-scale clinical trials with donezepil and the other cholinesterase inhibitors over periods of 15 and 24 weeks have yielded similar results for the three compounds in patients with mild to moderately severe AD. The drugs caused small but significant improvements in the ADAS-Cog, CIBIC, and MMSE scores. The most common side-effects were transient mild nausea, insomnia, and diarrhoea. Not all patients with AD will benefit from treatment with cholinesterase inhibitors; the proportion ranges from 30 to 50 per cent; although the clinical benefits of drug treatment in patients showing a response can persist for up to 24 months.

The approval of cholinesterase inhibitors for the treatment of Alzheimer's disease was an important landmark. They are reasonably well tolerated and produce significant, if modest, beneficial effects in patients with mild to moderately severe AD. However, they have not gained immediate and universal acceptance. In some countries (e.g. the United Kingdom) it has been argued that the drugs are too costly and provide at best only a modest improvement.

Some studies have also found the cholinesterase inhibitors to be effective in treating the cognitive deficits in vascular dementias, but the effects are small and less consistent. Rivastigmine has also been shown to have beneficial effects in treating the cognitive deficits in Parkinson's disease with dementia.[9] The cholinesterase inhibitors may thus find other applications as cognitive enhancers in conditions other than AD.

An alternative approach to cholinergic replacement therapy has been to develop **drugs that mimic acetylcholine** and act as agonists at the muscarinic cholinergic receptors in brain, but which, unlike acetylcholine itself, are bioavailable and brain-penetrant. Attention has focused on the discovery and development of muscarinic agonists that show selectivity for the m_1-receptor subtype, which is the predominant form present in the cerebral cortex. The most thoroughly studied cholinomimetic to date is xanomeline, a compound that acts as a highly potent and selective m_1-receptor agonist. Clinical effects were assessed in a multicentre study of 343 patients with AD.[10] Patients on the highest dose showed significant improvement when assessed using the ADAS-Cog scale and also showed a significant overall global improvement using CIBIC. In addition to cognitive improvements, patients receiving xanomeline also exhibited significant behavioural improvement, with dose-dependent reductions in vocal outbursts, suspiciousness, delusions, agitation, and hallucinations. Xanomeline is unlikely to be used in the treatment of AD because of its relatively short duration of action, but these results suggest that further research on cholinomimetics may still be justified.

An entirely different pharmacological approach is exemplified by the drug **memantine**, the first to be approved for the treatment of moderate to severe AD. Memantine is thought to act by virtue of its ability to block the NMDA sub-type of glutamate receptors in the brain. Clinical trials in AD showed small but significant beneficial effects on cognitive tests and in global clinical outcome, but curiously the effects were most notable in patients with advanced stage disease and less in patients with mild to moderate AD.[11] Although the effects of memantine are small, it remains the only effective treatment for advanced stage AD.

Attention-deficit hyperactivity disorder

Attention-deficit hyperactivity disorder (**ADHD**) is one of the most thoroughly studied disorders in child psychiatry, and the increasingly common use of stimulant drugs to treat this disorder has become the focus of much public attention and debate in recent years.[12] ADHD is defined in terms of three key features: lack of sustained attention, impulsivity, and hyperactivity. According to the DSM-IV definition the diagnosis of ADHD now includes more than 10 per cent of children.[13]

Because of the interest in the drug treatment of ADHD, a number of assessment tools have been developed. These include the widely used Conners' Teacher Rating Scale, the Conners' Parent Rating Scale, and a variety of tests designed to measure hyperactivity, problem behaviour, attention, and other aspects of cognition, as well as academic performance.[14]

The most commonly used drugs are the psychostimulants **amfetamine** and **methylphenidate**. Methylphenidate (Ritalin®) is by far the most widely prescribed. In more than 100 published trials these drugs have been found to have significant beneficial effects on all three key symptoms of ADHD in approximately 70 per cent of the treated children, and also in an adult form of ADHD.[15] Amfetamine (Adderall®) is approved for treatment of adult ADHD (not licensed in UK).

The mechanism of action of all three agents is similar; they act principally as inhibitors of the dopamine-uptake mechanism in the brain and promote the release of this neurotransmitter, thus stimulating dopaminergic mechanisms. The drugs also act to an important extent on noradrenaline-containing neurones to promote an increased release of this monoamine.[16] This may be relevant;a selective inhibitor of the noradrenaline transporter in brain, **atomoxetine**, has been approved as the first non-scheduled stimulant for the treatment of ADHD in both children and adults.[17]

Another non-amfetamine **modafinil** is also non-scheduled and is widely used in the United States for the treatment of ADHD,[18] although the drug has not yet gained regulatory approval from FDA because of possible serious adverse skin reactions. The mode of action of modafinil is unknown but it is used for the treatment of narcolepsy.

It is paradoxical that stimulant drugs, whose actions include an ability to promote hyperactivity, should have a calming effect

on hyperactive children. One explanation is that actions of these amfetamine derivatives show the 'rate dependency' typical of other central nervous system agents, i.e. they tend to stimulate low rates of behaviour and to suppress high rates.[19] An alternative view is that the relatively low doses of amphetamines used in the treatment of ADHD would not have stimulant effects even in normal healthy adults. There are few animal models that can be used in the study of psychostimulant use or ADHD. Mice that are genetically engineered to delete the genes for the dopamine and other monoamine transporters have proved valuable.[20] Animals which lack the dopamine transporter have elevated levels of dopamine in their brains and are behaviourally hyperactive. Paradoxically, d-amfetamine decreases activity in these animals, in keeping with the 'rate-dependency' hypothesis.[20]

The use of amfetamines, particularly methylphenidate, has increased rapidly during the past 30 years, particularly in the United States where in some states more than 10 per cent of school-age boys receive the drug.[12] The use of amfetamines in Europe has been at a much lower level so far, although their use in ADHD has also been increasingly rapidly.[12] In turn, such widespread use of psychostimulants creates problems about diversion and abuse.[12]

Further information

'Cognition Enhancers' in Drugs Futures 2025—[www.foresight.gov.uk/Brain_Science_Addiction].

'Drug Treatment of Alzheimer's Disease', Royal College of Psychiatry—[www.rcpsych.ac.uk/mentalhealthinformation/olderpeople/drugtreatmentofalzheimers.aspk].

'Methylphenidate, atomoxetine and dexamfetamine for attention deficit hyperactivity disorder (ADHD) in children and adolescents—guidelines.' National Institute for Clinical Excellence UK, March 22, 2006. [http://www.nice.org.uk].

References

1. Moreira, P.I., Zhu, X., Nunomura, A., *et al.* (2006). Therapeutic options in Alzheimer's disease. *Expert Review of Neurotherapeutics*, **6**, 897–910.
2. Gershon, S., Ferris, S.H., Kennedy, J.S., *et al.* (1994). Methods for the evaluation of pharmacological agents in the treatment of cognitive and other deficits in dementia. In *Clinical evaluation of psychotropic drugs: principles and guidelines* (eds. R.F. Prien and D.S. Robinson), pp. 467–99. Raven Press, New York.
3. Bartus, R.T., Dean, R.L., Beer, B., *et al.* (1982). The cholinergic hypothesis of geriatric memory dysfunction. *Science*, **217**, 408–17.
4. Rupniak, N.M.J., Steventon, M.J., Jennings, C.A., *et al.* (1989). Comparison of the effects of four cholinomimetic agents on cognition in primates following disruption by scopolamine. *Psychopharmacology*, **99**, 189–95.
5. Giacobini, E. (1998). Cholinesterase inhibitors for Alzheimer's disease therapy: from tacrine to future applications. *Neurochemistry International*, **32**, 413–19.
6. Birks, J. and Harvey, R.J. (2006). Donepezil for dementia due to Alzheimer's disease. *Cochrane Database Systematic Reviews*, (1), CD001190.
7. Loy, C. and Schneider, L. (2004). Galantamine for Alzheimer's disease. *Cochrane Database Systematic Reviews*, (4), CD001747.
8. Birks, J., Grimley Evans, J., Lakovidou, V., *et al.* (2006). Rivastigmine for Alzheimer's disease. *Cochrane Database Systematic Reviews*, (4), CD001191.
9. Maidment, I., Fox, C., and Boustani, M. (2006). Cholinesterase inhibitors for Parkinson's disease dementia. *Cochrane Database Systematic Reviews*, (1), CD004747.
10. Bodick, N.C., Offen, W.W., Levey, A.I., *et al.* (1997). Effects of xanomeline, a selective muscarinic receptor agonist, on cognitive function and behavioral symptom in Alzheimer's disease. *Archives of Neurology*, **54**, 465–73.
11. McShane, R., Areosa Sastre, A., and Minakaran, N. (2006). Memantine for dementia. *Cochrane Database Systematic Reviews*, (2), CD003154.
12. Iversen, L.L. (2006). *Speed, ecstasy, ritalin: the science of amphetamines*. Oxford University Press, Oxford.
13. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, Text Revision) (DSM-IV-TR). American Psychiatric Association, Washington, DC.
14. Conners, C.K. (1998). Rating scales in attention deficit hyperactivity disorder: use in assessment and treatment monitoring. *The Journal of Clinical Psychiatry*, **59**(Suppl. 7), 24–30.
15. Wender, P.H. (2001). *ADHD: attention-deficit hyperactivity disorder in children, adolescents and adults*. Oxford University Press, Oxford.
16. Solanto, M.V. (1998). Neuropsychopharmacological mechanism of stimulant drug action in attention-deficit hyperactivity disorder: a review and integration. *Behavioural Brain Research*, **94**, 127–52.
17. Kratochvil, C.J., Vaughan, B.S., Daughton, J.M., *et al.* (2004). Atomoxetine in the treatment of attention deficit hyperactivity disorder. *Expert Review of Neurotherapeutics*, **4**, 601–11.
18. White, R.F. and Giorgadze, A. (2006). A randomized, double-blind, placebo-controlled study of modafinil film-coated tablets in children and adolescents with attention-deficit hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**, 503–11.
19. Robbins, T.W. and Sahakian, B.J. (1979). Paradoxical effects of psychomotor stimulant drugs in hyperactive children from the standpoint of behavioural pharmacology. *Neuropharmacology*, **18**, 931–50.
20. Gainetdinov, R.R., Sotnikova, T.D., and Caron, M.G. (2002). Monoamine transporter pharmacology and mutant mice. *Trends in Pharmacological Sciences*, **23**, 367–73.

6.2.8 Drugs used in the treatment of the addictions

Fergus D. Law and David J. Nutt

Medical treatment of the addictions remains controversial, with addiction itself viewed as a lifestyle problem, a hijacking of brain systems by drugs, or as a medical illness. Many of these controversies may be avoided by taking a goal-oriented approach to treatment, in which clinical objectives are defined, and both medications and psychological interventions are used to facilitate progress towards these. The effectiveness of medications is maximized when they are used as one component of a comprehensive treatment plan.

There are no 'magic bullets' in addiction treatment—the same pharmacological principles apply to these drug treatments as to any other. Drugs need to be given in effective doses, at appropriate intervals, allowed time to reach steady state, and also to dissipate when terminated on the basis of their half-life. Some drugs also have an abuse potential of their own (e.g. opiates, sedative-hypnotics) especially those with a rapid onset of action, and such

drugs need to be particularly closely monitored and controlled, to minimize their diversion and misuse.

Medications in perspective

The clinical goal-oriented approach requires clarity about the clinical objectives at each phase of the treatment process. A typical treatment plan involves three primary clinical objectives:

- drug and psychosocial stabilization
- detoxification when appropriate
- prevention of relapse or recurrence.

Stabilization with a substitution treatment (e.g. methadone or buprenorphine in opiate addiction) involves prescribing a pharmacological equivalent to the abused drug to stop illicit use, crime, etc. It allows time for stabilization to occur and to consider later objectives. Stabilization itself may be either short-term with the primary goal of terminating illicit drug use 'on top' of the prescription, or longer-term where it is commonly known as maintenance. The goals of maintenance are either psychosocial stabilization in preparation for detoxification, or harm reduction in patients where abstinence is not practicable or safe. Unless a sufficient degree of stabilization has been achieved prior to detoxification, the chances of success are strictly limited. The harm-reduction goal has generated much controversy, but one of its major benefits is the reduction in the spread of HIV among injecting drug users. In this group stable long-term maintenance treatment is preferable to repeated cycles of premature discontinuation followed by relapse to uncontrolled drug use with its attendant elevated risk of death from overdose as tolerance wanes. Monitoring by using drug screens and clinical assessments is required to ensure that patients do not use 'on top' of their prescription.

Specific drugs used in addiction treatment

This chapter deals with methadone, levacetylmethadol (LAAM), codeine and dihydrocodeine tartrate, buprenorphine, clonidine, lofexidine, naltrexone, naloxone, acamprosate, disulfiram, and clomethiazole (chlormethiazole), and covers addiction indications only. Many of these drugs are not licensed for use in addiction treatment (e.g. methadone tablets and injection), or are currently licensed in only one or a few countries (e.g. clonidine in Germany, lofexidine in the United Kingdom).

Methadone[1,2] (Methadose®, Physeptone®, Synastone®)

This is a long-acting opioid analgesic which has been the mainstay of opioid substitution treatment, but is often difficult to stop due to its prolonged withdrawal syndrome.

Pharmacology: it is a strong full μ-opioid agonist.

Types of compounds available: it is available in liquid, injectable, and tablet formulations.

Pharmacokinetics: t_{max} 2 to 4 h after oral dosing; 1 h after intramuscular injection; its half-life is 25 h.

Side-effects: as with other μ-opioids, its side-effects include mental blunting, sweating, constipation, nausea, and analgesia.

Toxic effects: acute overdose leads to respiratory depression and pulmonary oedema.

Indications: opiate maintenance, stabilization, and detoxification; it is also indicated for use during pregnancy.

Contraindications: respiratory or severe liver disease, monoamine oxidase inhibitors; caution should be exercised in elderly people.

Interactions: respiratory depression especially in combination with other sedative drugs; metabolism affected by hepatic enzyme induction and inhibition; plasma levels are affected by HAART drugs.

Effects of withdrawal: these include moderate but prolonged abstinence syndrome, especially poor sleep.

Dosage and administration: single daily dose, occasionally twice daily; dose depends on the level of dependence—if unknown, initially 10 to 20 mg daily. The minimum dose that covers withdrawal symptoms for 24 h should be given, and increased by 5 to 10 mg as necessary. Close monitoring is necessary by clinical assessment and drug screen. Some centres use intravenous preparations in those who don't respond to oral preparations.

LAAM[3] (ORLAAM®)

LAAM is a methadone variant with a much longer half-life, requiring only three visits a week for full supervision of medication. However, take-home medication is not allowed and its use is restricted to specialist clinics and is reserved for patients who have failed other treatments. Its licence in Europe has been withdrawn due to QTc prolongation.

Pharmacology: it is a synthetic μ-opioid agonist with active metabolites which are more potent than the parent drug.

Types of compounds available: aqueous solution.

Pharmacokinetics: t_{max} 2 to 4 h; duration of action is 48 to 72 h; half-lives for LAAM and its metabolites are 2 to 4 days; it takes 2 weeks to reach steady state.

Side-effects: too rapid escalation of the dose may result in sedation, orthostatic hypotension, poor concentration, and overdose.

Toxic effects: as for methadone; QTc prolongation; overdose occurs with too frequent (daily) dosing, use of multiple drugs, or 'on-top' use due to impatience with its slow onset of action.

Indications: opiate maintenance, stabilization, and detoxification.

Contraindications: pregnancy (transfer to methadone); QTc prolongation prior to induction of treatment; dose should be reduced in elderly people, and in renal and hepatic impairment.

Interactions: as for methadone; other drugs prolonging QTc interval.

Effects of withdrawal: as for methadone, but with a milder withdrawal syndrome due to longer $t_{1/2}$.

Dosage and administration: pre-treatment ECG to identify prolonged QTc intervals, repeated 12–14 days after initiating treatment and periodically thereafter to rule out alterations to the QTc; give three times a week or on alternate days; increase dose by 20 to 40 per cent when transferring form 48 h to 72 h dosing interval. Transfer methadone to LAAM by giving 1.2 to 1.3 times the daily methadone dose; and LAAM to methadone by waiting at least 48 h and then giving 0.8 times the LAAM dose. If low or unknown tolerance, the initial dose is 20 to 40 mg three times weekly. Adjust dose in 5- to 10-mg steps, but no more frequently than

weekly at the most. Strongly warn patients of the risk of supplementation with street drugs especially prior to steady state. LAAM is detected by urine screens for methadone.

Buprenorphine[4,5] (Subutex®, Suboxone®, Temgesic®, Buprenex®)

Advantages over methadone and other full μ-agonists are its safety in overdose, the attenuation of the drug 'high' during on-top use, and its low levels of psychological reinforcement and withdrawal symptomatology during detoxification.

Pharmacology: a partial μ-opioid agonist, which explains the ceiling on respiratory depression; slow onset of action; dissociates slowly from the μ-receptor.

Types of compounds available: 0.2-, 0.4-, 2- or 8-mg sublingual tablets, or 0.3-mg ampoules for injection; it is also available in combination with naloxone in a 1:4 ratio (Suboxone®) to reduce misuse if diverted.

Pharmacokinetics: sublingual tablets absorbed rapidly into the buccal mucosa and released slowly into the blood stream; t_{max} 2 to 6 h.

Side-effects: withdrawal symptoms if either too little or too much is given; nausea and vomiting are rare in addicts.

Toxic effects: as for methadone, but less constipation and respiratory depression.

Indications: opiate maintenance, stabilization, and detoxification, including in pregnancy; may be especially suitable for opioid antagonist-assisted withdrawal; no dosage adjustment needed in renal failure or elderly people.

Contraindications: severe respiratory disease; use with care in severe liver disease.

Interactions: rare; sedation with benzodiazepines.

Effects of withdrawal: there is a mild but delayed withdrawal syndrome.

Dosage and administration: initial dose is 0.8–4 mg increasing by 4 to 8 mg daily until the required dose level is reached; usual daily dose 8 to 32 mg; doses above 12 mg may be given on alternate days; minimize withdrawal symptoms after long-term use by reducing by 1 mg every 3 to 4 days or less often; buprenorphine-assisted heroin detoxification by rapid reduction over 5 to 10 days; the injectable form is not recommended for use in addiction treatment. Monitor clinical state and perform drug screens for compliance and on-top use.

Codeine phosphate and dihydrocodeine tartrate[6,7] (DF118 Forte®, DHC Continus®)

Advantages over methadone occur in situations where long-acting opioids may be inappropriate. These are often preferred by patients, and by doctors treating the young, low-dose users, and in acute situations (e.g. in police custody). Disadvantages are its ease of misuse, high levels of psychological reinforcement due to its rapid onset of action, and its unfavourable side-effect profile.

Pharmacology: it is a weak short-acting μ-opioid agonist.

Types of compounds available: Codeine: 15-, 30- and 60-mg oral tablets, linctus, syrup and injection; dihydrocodeine tartrate: 30- or 40-mg oral tablets for use three to six times a day, and a 60-, 90-, or 120-mg slow-release preparation (DHC Continus) for use every 12 h; also parenteral preparation and elixir.

Pharmacokinetics: peak plasma levels at 1 to 2 h; half-life 3.5 to 4.5 h.

Side-effects: it is more likely to cause sedation, dizziness, stimulation, euphoria, constipation, histamine release, psychomimetic effects, and disturbing dreams than other opioids.

Toxic effects: precipitation of life-threatening exacerbations of asthma; in overdose, coma with myotonic twitching, grand mal convulsions, and rarely rhabdomyolysis may occur.

Indications: opiate maintenance, stabilization, and detoxification

Contraindications: acute exacerbations of asthma, lower respiratory tract infection, respiratory depression, and hepatic failure, increased intracranial pressure; caution should be exercised in renal impairment and elderly people.

Interactions: as for methadone; it may enhance the effects of warfarin.

Effects of withdrawal: mild withdrawal syndrome.

Dosage and administration: dose depends on level of dependence; monitor clinical state and urine screen to confirm compliance and termination of on-top use.

Naltrexone[8–11] (Nalorex®, Opizone®)

Naltrexone is used to maintain abstinence in detoxified opiate addicts during the period of highest vulnerability to relapse following detoxification. It blocks the 'high' produced by opiates and promotes the extinction of conditioned responses. It also has a role in alcohol misuse and ultra-rapid opioid detoxification. Nalmefene is a related long-acting μ receptor antagonist that has a licence for alcoholism in some countries.

Pharmacology: it is a long-acting non-selective opioid antagonist.

Type of compound available: 50-mg oral tablet.

Pharmacokinetics: t_{max} 1 h; duration of action is dose related, and a single dose can be effective for up to 48 h.

Side-effects: opiate withdrawal syndrome may occur on induction; occasionally, gastrointestinal irritation, headaches, arthralgia, flattening of mood, and rash occur.

Toxic effects: severe opioid withdrawal in dependent addicts lasting 2 days; reversible liver toxicity at high dose in obese and elderly people; liver function tests should be monitored especially if baseline tests are impaired.

Indications: prevention of impulsive relapse following detoxification in opioid users; opioid antagonist-assisted withdrawal; it reduces the reinforcing effects of alcohol.

Contraindications: acute hepatitis or liver failure, and active peptic ulcer; caution should be exercised in hepatic or renal impairment.

Interactions: competitive opioid blockade, so potentially can be overcome using very high opiate doses.

Effects of withdrawal: none, but risk of opioid overdoes following withdrawal (loss of toerance).

Dosage and administration: treatment initiated following LFTs and opioid-negative urine screen (or a negative naloxone challenge). Twenty-five mg is given on the first day, and then 50 mg daily for 3 to 6 months. Thrice weekly dosing (100/100/150 mg) may occasionally improve compliance. Supervision of consumption by a supportive person, urine tests to monitor compliance, and regular reviews are very important in maximizing effectiveness.

Naloxone[12] (Narcan®)

Naloxone is a short-acting antagonist used in the treatment of opioid overdose, during detoxification, and naltrexone induction. Take home naloxone may be used to treat opiate overdoses in the community by patients trained in its use.

Pharmacology: it is a short-acting competitive opioid antagonist.

Types of compounds available: it is available in injectable form for intramuscular, intravenous, or subcutaneous use.

Pharmacokinetics: half-life 1 to 2 h.

Side-effects: withdrawal in opiate-dependent subjects; nausea and vomiting; rarely, high blood pressure and pulmonary oedema can occur.

Toxic effects: very occasional deaths due to acute pulmonary oedema, extreme hypertension, and ventricular arrhythmias have occurred in those with known myocardial disease.

Indications: naloxone reverses the effects of opioid overdose, and in high doses may help in overdose due to alcohol and benzodiazepines. Naloxone is also occasionally used as the diagnostic test of opioid dependence, and as a challenge test prior to naltrexone initiation. It is also used for opioid antagonist-assisted withdrawal, and in combination with oral opiate agonists to reduce misuse by the injectable route (e.g. Suboxone®). In neonates it is used for the reversal of the effects of opioids given to mothers during labour.

Contraindications: none if not opioid dependent (safe in neonates, children, pregnancy, elderly people); caution is advisable in opioid dependence, painful conditions, and cardiovascular disease.

Interactions: severe hypertension following reversal of coma due to clonidine overdose.

Effects of withdrawal: none.

Dosage and administration: in opioid overdose, give 0.4 to 2 mg intravenously (5 to 10 μg/kg in neonates and children) and repeat at 2- to 3-min intervals until desired response; may also be given intramuscularly in overdose; lower doses are used in adults (0.1–0.2 mg) to reverse opioid-induced respiratory depression, but higher doses are needed with buprenorphine and σ-receptor agonists. Continue naloxone by infusion or repeated injection if necessary to maintain recovery. During naltrexone induction, give 0.2 mg parenterally followed by 0.8 mg 30 min later (or 0.6 mg 30 s later if given intravenously). In equivocal cases give 1.6 mg.

Clonidine hydrochloride[13, 14] (Catapres®, Dixarit®)

Clonidine is an α_2-adrenoceptor agonist used to suppress some symptoms of opioid withdrawal, especially methadone-assisted withdrawal. It is ineffective for subjective symptoms, muscle/bone aches, stomach cramps, and insomnia.

Pharmacology: clonidine is an antihypertensive agent that decreases central and peripheral (sympathetic) noradrenergic activity by stimulating presynaptic receptors in the locus coeruleus.

Types of compounds available: it is available as tablet, liquid, sustained release capsule, and transdermal preparation; the tablet is licensed in Germany.

Pharmacokinetics: t_{max} 90 min; half-life of 20 to 25 h

Side-effects: it causes hypotension, sedation, dry mucous membranes, bradycardia, depression, impotence, constipation and diarrhoea, sleep disturbance, fluid retention, headache, euphoria, and Raynaud's phenomenon.

Toxic effects: a clonidine withdrawal syndrome may occur on abrupt withdrawal or non-compliance; paralytic ileus, psychotic features, or depression can also occur; coma or severe sedation can occur on acute overdose.

Indications: rapid opiate withdrawal and opiate antagonist-assisted withdrawal; it is also used as adjunct for alcohol, benzodiazepine, and nicotine withdrawal.

Contraindications: low baseline blood pressure, disorders of cardiac pacemaker activity and conduction, cardiovascular and cerebrovascular disease, and porphyria; caution is necessary in renal and hepatic impairment, peripheral vascular disease, and where there is a history of depression or psychosis.

Interactions: combinations with sedative drugs; phenothiazines may increase hypotension. Tricyclic antidepressants may block its effects.

Effects of withdrawal: an increase in sympathetic activity with symptoms mimicking the opiate withdrawal syndrome may occur 18 to 72 h after the last dose on abrupt termination. Blood pressure rebound is rare when used for less than 1 month. The effects are minimized by gradual withdrawal.

Dosage and administration: expertise is needed to monitor cardiovascular signs and adjust dose during a clonidine detoxification over 1 to 3 weeks. Start clonidine after discontinuation of the opioid. Following a test dose, 0.1 mg tablets are given four to six times daily building up to 2 mg daily over a few days in inpatients but half this dose in outpatients. Patches applied once weekly and supplemented by tablets if withdrawal symptoms occur. Frequent monitoring for hypotension and bradycardia is needed.

Lofexidine[13–15] (BritLofex®)

Lofexidine is an analogue of clonidine, but easier to use because there is less hypotension and sedation.

Pharmacology: as for clonidine; differences occur possibly because it is more potent at the A subtype of α_2-adrenoceptors.

Types of compounds available: 0.2-mg oral tablets; these are licensed in the United Kingdom only.

Pharmacokinetics: t_{max} 3 h; half-life of 15 h.

Side-effects: these are the same as for clonidine, but with markedly less hypotension and other side-effects.

Toxic effects: as for clonidine; it has little rebound effect on blood pressure, and no psychiatric complications or misuse has been reported.

Indications and contraindications: as for clonidine.

Interactions: as for clonidine.

Effects of withdrawal: as for clonidine, but less rebound.

Dosage and administration: treat for 1 to 3 weeks; dose may be started before opiate is stopped; on day 1, give 0.8 mg in divided doses, and build up by 0.4 to 0.8 mg daily. Aim for a minimum of 1.6 mg daily in four divided doses, increasing to a maximum daily dose of 2.4 mg. Plan for a peak in the dose when the peak of withdrawal symptoms are expected. Blood pressure should be monitored 2 h after the initial dose, and daily as the dose is increasing. After peak opiate withdrawal lofexidine should be withdrawn gradually over at least 2–4 days. The dose should be reduced by 0.2–0.4 mg daily.

Acamprosate (calcium acetylhomotaurinate)[16,17] (Campral EC®)

Acamprosate is a non-aversive agent used to maintain abstinence in alcohol-dependent patients during the most vulnerable period following detoxification, which may work as an anticraving agent.

Pharmacology: this is not fully understood. It affects the brain's γ-aminobutyric acid (inhibitory) and glutamate (excitatory) systems.

Types of compounds available: oral enteric-coated tablets containing 333 mg acamprosate.

Pharmacokinetics: t_{max} 5 h; half-life 21 h; steady state after 7 days.

Side-effects: there are a range of dose-related side-effects which are mainly mild and transient, including diarrhoea and other gastrointestinal effects, pins and needles in the limbs, skin pruritus, confusion, and sexual effects. Transient reductions in blood pressure occur in those with alcohol-induced hepatic cirrhosis.

Toxic effects: hypercalcaemia is a theoretical possibility following overdose.

Indications: maintenance of abstinence following alcohol detoxification; it is suitable for use in those with liver dysfunction.

Contraindications: renal impairment and severe hepatic failure.

Interactions: concomitant food decreases bioavailability.

Effects of withdrawal: none.

Dosage and administration: begin treatment as soon as possible after detoxification, and continue for 1 year. Four tablets a day (2–1–1) with meals if body weight is less than 60 kg, but six tablets a day (2–2–2) if over 60 kg. The drug should be continued during alcohol relapses.

Disulfiram[18–20] (Antabuse®)

An unpleasant disulfiram-ethanol reaction occurs when alcohol is consumed, acting as a deterrent or punishment if drinking occurs. Disulfiram is used under specialist supervision during periods of vulnerability to relapse.

Pharmacology: disulfiram is an aldehyde dehydrogenase inhibitor leading to the accumulation of acetaldehyde after ethanol consumption.

Types of compounds available: 200-mg oral tablets.

Pharmacokinetics: inhibition of alcohol dehydrogenase develops slowly over 12 to 24 h and peaks at 48 h.

Side-effects: relatively non-toxic on its own, but may cause drowsiness, fatigue, halitosis, nausea, vomiting, and a decrease in libido. With alcohol, disulfiram causes nausea, vertigo, anxiety, blurred vision, hypotension, chest pain, palpitations, tachycardia, facial flushing, and throbbing headache. Symptoms can last 3 to 4 days, but may persist for 1 week. Symptoms may occur even with small amounts of alcohol, but 25 to 50 per cent of patients experience little or no reaction at standard doses.

Toxic effects: the disulfiram-ethanol reaction may be very severe with respiratory depression, cardiovascular collapse, cardiac arrhythmias, coma, cerebral oedema, hemiplegia, convulsions, and death. Chronic treatment and overdose may cause high blood pressure, hepatotoxicity, and neuropsychiatric complications.

Indications: it is used as a deterrent to the use of alcohol and maintenance of abstinence, especially if there is high motivation and good compliance in the patient.

Contraindications: cardiac failure, cardiovascular or cerebrovascular disease, hypertension, peripheral neuropathy, psychosis, severe personality disorder, suicide risk, pregnancy, and breast feeding. It should be used with caution in hypertension, diabetes mellitus, epilepsy, impaired hepatic or renal function, respiratory disorders, cerebral damage, and hypothyroidism, as it may exacerbate these conditions. Caution should be exercised in the elderly.

Interactions: caution with phenytoin, diazepam, chlordiazepoxide, theophylline, warfarin, and caffeine metabolism; acute psychosis or confusional state may occur with metronidazole. Concurrent tricyclic antidepressants may exacerbate the disulfiram-ethanol reaction and cause a toxic confusional state.

Effects of withdrawal: none but restoration of alcohol dehydrogenase depends on *de novo* enzyme synthesis which occurs over 6 or more days.

Dosage and administration: after 24 h without alcohol, give 800 mg as a single dose on day 1, then reduce dose over 5 days from 100 to 200 mg daily. Effectiveness is dose related. Blood pressure should be monitored regularly if the patient is taking over 500 mg/day. Compliance is improved with monitoring (carbon disulphide breath test) and supervision. Patients should be warned not to ingest any alcohol, including alcohol in food, liquid medicines, and even toiletries. An alcohol challenge test may be done in specialist centres. The patient should be reviewed every 6 months at a minimum. Alcohol should be avoided for at least 1 week on terminating disulfiram.

Clomethiazole (edisylate)[21] (Heminevrin®)

Clomethiazole (previously known as chlormethiazole) is a sedative-hypnotic-anxiolytic which also inhibits the metabolism of alcohol resulting in a more gradual elimination of alcohol from the body.

Pharmacology: it is an agonist at the picrotoxin/barbiturate site of the GABA-A receptor, a glutamate antagonist, and an inhibitor of alcohol dehydrogenase.

Types of compounds available: oral and parenteral forms.

Pharmacokinetics: t_{max} 1 h (oral dosing); plasma half-life is 4 h but is double this in elderly people.

Side-effects: conjunctival irritation, nasal congestion, tingling in the nose, headaches, and reversible elevation of liver function tests.

Toxic effects: respiratory depression, sudden fall in blood pressure, anaphylactic reactions, and death (often due to combination with alcohol).

Indications: acute alcohol withdrawal and delirium tremens in inpatients.

Contraindications: alcohol addicts who continue to drink, pulmonary insufficiency, pregnancy, and lactation; caution is advised in renal impairment, severe liver damage, and cardiac and respiratory disease.

Interactions: alcohol and diazoxide may cause severe respiratory depression. Plasma levels are increased by cimetidine. It causes severe bradycardia with propranolol.

Effects of withdrawal: rebound insomnia and anxiety (as with other sedative drugs).

Dosage and administration: titrate using three or four daily doses according to patient response. Initially give 2 to 4 capsules, then 9 to 12 in divided doses over the next 24 h, 6 to 8 capsules on

day 2, 4 to 6 on day 3, reducing it to 0 by day 6 to 9 in order to avoid dependency. Only use it by infusion where resuscitation facilities are available; initially give 3 to 7.5 ml/min, then reduce dosage to 0.5 to 1 ml/min (infusion no longer available in UK).

Further information

British National Formulary (BNF) www.bnf.org.
NIDA (National Institute on Drug Abuse) www.nida.nih.gov.
SAMHSA (Substance Abuse and mental health services administration). www.samhsa.gov

References

1. Lingford-Hughes, A.R., Welch, S., and Nutt, D.J. (2004). Evidence-based guidelines for the pharmacological management of substance misuse, addiction and comorbidity: recommendations from the British Association for Psychopharmacology. *Journal of Psychopharmacology*, **18**, 293–335.
2. NICE. (2007). *Methadone and buprenorphine for the management of opioid dependence. NICE technology appraisal guidance 114*. National Institute for Health and Clinical Excellence, London.
3. Clark, N., Lintzeris, N., Gijsbers, A., *et al.* (2002). LAAM maintenance vs methadone maintenance for heroin dependence. *The Cochrane Database of Systematic Reviews*, (2): CD002210.
4. Law, F.D., Myles, J.S., Daglish, M.R.C., *et al.* (2004). The clinical use of buprenorphine in opiate addiction: evidence and practice. *Acta Neuropsychiatrica*, **16**, 246–74 (with erratum at 2004, **16**, 326).
5. Nutt, D.J. (1997). Receptor pharmacology of buprenorphine. *Research and Clinical Forums*, **19**, 9–15.
6. Krausz, M., Verthein, U., Degwitz, P., *et al.* (1998). Maintenance treatment of opiate addicts in Germany with medications containing codeine—results of a follow up study. *Addiction*, **93**, 1161–7.
7. Robertson, J.R., Raab, G.M., Bruce, M., *et al.* (2006). Addressing the efficacy of dihydrocodeine versus methadone as an alternative maintenance treatment for opiate dependence: a randomized controlled trial. *Addiction*, **101**, 1752–9.
8. Gonzalez, J.P. and Brogden, R.N. (1988). Naltrexone. A review of its pharmacodynamic and pharmacokinetic properties and therapeutic efficacy in the management of opioid dependence. *Drugs*, **35**, 192–213.
9. NICE. (2007). *Naltrexone for the management of opioid dependence. NICE technology appraisal guidance 115*. National Institute for Health and Clinical Excellence, London.
10. Srisurapanont, M. and Jarusuraisin, N. (2005). Opioid antagonists for alcohol dependence. *The Cochrane Database of Systematic Reviews*, (1): CD001867.
11. Minozzi, S., Amato, L., Vecchi, S., *et al.* (2006). Oral naltrexone maintenance treatment for opioid dependence. *The Cochrane Database of Systematic Reviews*, (1): CD001333.
12. Handal, K.A., Schauben, J.L., and Salamone F.R. (1983). Naloxone. *Annals of Emergency Medicine*, **12**, 438–45.
13. Gowing, L., Farrell, M., Ali, R., *et al.* (2004). Alpha2 adrenergic agonists for the management of opioid withdrawal. *The Cochrane Database of Systematic Reviews*, (4): CD002024.
14. Kahn, A., Mumford, J.P., Rogers, GA., *et al.* (1997). Double-blind study of lofexidine and clonidine in the detoxification of opiate addicts in hospital. *Drug and Alcohol Dependence*, **44**, 57–61.
15. Strang, J., Bearn, J., and Gossop, M. (1999). Lofexidine for opiate detoxification: a review of recent randomised and open controlled trials. *American Journal on Addictions*, **8**, 337–48.
16. Littleton, J. (1995). Acamprosate in alcohol dependence: how does it work? *Addiction*, **90**, 1179–88.
17. Wilde, M.I. and Wagstaff, A.J. (1997). Acamprosate: a review of its pharmacology and clinical potential in the management of alcohol dependence after detoxification. *Drugs*, **53**, 1038–53.
18. Banys, P. (1988). The clinical use of disulfiram (Antabuse): a review. *Journal of Psychoactive Drugs*, **20**, 243–60.
19. Heather, N. (1993). Disulfiram treatment for alcohol problems: is it effective and, if so, why? In *Treatment options in addiction: medical management of alcohol and opiate abuse* (ed. C. Brewer), pp. 1–18. Gaskell, London.
20. Brewer, C. (1984). How effective is the standard dose of disulfiram? A review of the alcohol-disulfiram reaction in practice. *The British Journal of Psychiatry*, **144**, 200–2.
21. Majumdar, S.K. (1990). Chlormethiazole: current status in the treatment of the acute ethanol withdrawal syndrome. *Drug and Alcohol Dependence*, **27**, 201–7.

6.2.9 Complementary medicines

Ursula Werneke

Complementary medicines pose a particular challenge to medical practitioners who may feel that their patients need conventional treatment but often find themselves out of their depth when patients ask about complementary therapies. Pharmacological options include herbal medicines, certain foods, and nutritional supplements such as vitamins and minerals. Physical treatments include acupuncture, massage, and osteopathy to name a few. Treatments, which purport to achieve their effects through changes in internal 'energy flow' include reiki, reflexology, healing, and therapeutic touch, and also homeopathy and traditional Chinese acupuncture. All these treatments are either used alternatively, i.e. instead of, or complementary, i.e. in addition to, conventional medicine. In patients with mental health problems, depending on the definition and inclusion criteria, estimates of the prevalence of complementary medicine use range from 8 per cent to 57 per cent. Depression and anxiety seem to be the most common indications.[1]

Herbal remedies and supplements

Principles of treatment

Herbal remedies and supplements may come in many different forms and formulations. Since they are currently not subject to the same regulatory requirements as conventional medicines they can vary substantially in contents and dose even if they purport to contain the same ingredients. The pharmacological properties of an extract or a supplement may depend on many different factors (Table 6.2.9.1).

Condition-specific remedies

Cognitive enhancers

Cognitive enhancers are either used in the treatment of dementia to enhance mental performance or prevent cognitive decline in healthy people. One strategy aims at increasing choline availability, e.g. by inhibiting acetylcholine esterase. Alternative non-cholinergic neuroprotective strategies have been postulated. These rely on antioxidants scavenging free radicals thereby reducing neurotoxicity or anti-coagulants and increasing cerebral blood flow.[4] Suggested herbal cognitive enhancers for which some positive trial evidence has been collated include ginkgo (*Ginkgo biloba*), panax

Table 6.2.9.1 Determinants of pharmacological properties of complementary medicines

Factor	Problem	Example
Material production	Quality may depend on plant material used, time of harvesting, geographical location, or other environmental factors	St John's wort extracts prepared from the flowers are more potent than extracts prepared from the leaves
Extraction method	Determines remedy composition	Alcoholic valerian extracts may be safer than aqueous extracts such as teas because harmful volatile substances (valepotriates) are eliminated more easily. The resulting extracts may be less potent though, since valepotriates have GABAergic properties.[2] Conversely, aqueous kava extracts may be safer than alcoholic extracts because liver protective substances such as glutathione are retained
Standardization	Difficult to achieve if active ingredient is unknown	St John's wort is based on the extract traditionally standardized on hypericin, a photosensitive red pigment. However, current evidence suggests that standardization should be based on hyperforin, which inhibits the reuptake of monoamines[1]
Dosing	Depends on standardization	300 mg of St John's wort extract standardized on 0.5% or 5% hyperforin most likely have different pharmacodynamic effects
Contamination	Increased and sometimes unexpected toxicity	Contamination with e.g. fertilizer residuals or heavy metals. Association of eosinophilicmyalgic syndrome and some 5-hydroxy-tryptophan products may be at least in part due to contamination of some batches
Adulteration	May lead to serious side effects and drug interactions falsely ascribed to the remedy *per se*	Adulterants include steroids, NSAIDs, anticonvulsants, benzodiazepines, hypoglycemic agents, erectogenic agents, and warfarin[3]

ginseng (*Panax ginseng*), hydergine (*Claviceps purpurea*), sage (*Salvia officinalis*), and vitamin E. The potential side effects can be derived from the purported mechanisms of action. For example, remedies increasing the cerebral blood flow such as ginkgo may increase the risk of cerebral haemorrhage. *Panax ginseng* has been associated with manic episodes and hydergine can lead to ergot poisoning unless dosed carefully. Some sage species can lower the seizure threshold. Others contain camphor, which can be toxic in high doses.[2] Also, evidence is emerging that taking vitamin E above the recommended level may increase all-cause mortality.[5]

Anxiolytics and sedatives

Drugs considered to be either anxiolytics or sedatives essentially have the same underlying mechanisms of action. The stronger an agent the more sedating it will be, leading to coma in extreme cases. Four mechanisms of action have been implicated; binding to γ-aminobutyric acid (GABA) receptors leading to hyperpolarization of the cell membrane through increased influx of chlorine anions; inhibition of excitatory amino acids (EAA) thereby also impairing the ability to form new memories; sodium channel blockade, reducing depolarization of the cell membrane; and calcium channel blockade, reducing the release of neurotransmitters into the synaptic cleft.[4] The most commonly used CAMs for anxiolysis and sedation, such as valerian (*Valeriana officinalis*), passion flower (*Passiflora incarnata*), kava (*Piper methysticum*), and German chamomile (*Matricaria recutita*) are GABAergic. Lemon balm (*Melissa officinalis*) has cholinergic and GABAergic properties. For other plant remedies, including hops (*Humulus lupulus*), oats (*Avena sativa*), lavender (*Lavendula angustifolia*), and starflower also known as borage (*Borago officinalis*), the actual mechanism of action remains unknown. Melatonin regulates the circadian rhythm and also has some GABAergic properties although trial evidence remains inconclusive. Some of these remedies can potentially have serious side effects. For instance, kava extracts have been associated with significant and potentially fatal hepatotoxicity. Starflower contains γ-linolenic acid that may lower the seizure threshold. Some passion flower extracts may contain cyanide components. As expected, all remedies in this class can lead to drowsiness when taken in high doses and can potentiate the effect of synthetic sedatives.

Antidepressants and mood stabilizers

Most complementary antidepressants are thought to work through serotonergic and noradrenergic pathways. The most robust clinical data are available for St John's wort (*Hypericum perforatum*), having been extensively reviewed in meta-analyses.[1] Hyperforin, inhibiting the reuptake of monoamines, is thought to be the most likely active component. Supplements, such as *S*-adenosylmethionine (SAMe), folic acid, L-tryptophan, and 5-hydroxytryptophan are components or co-factors in the serotonin synthesis. For SAMe, equivalence to tricyclic antidepressants has been demonstrated. However, SAMe is very expensive and a suitable oral formulation may be difficult to obtain. Selenium has also been suggested but the mechanism of action, albeit still unclear, seems to be different. Its antioxidant properties may reduce nerve cell damage. Selenium also facilitates conversion from thyroxin (T4) to thyronine (T3), and T3 substitution is one possible augmentation strategy for antidepressants. As for lithium, the therapeutic index is narrow. Omega-3 fatty acids are known to stabilize membranes and to facilitate monoaminergic, serotonergic, and cholinergic neurotransmission. The currently available evidence supports the use of eicosapentaeonic acid on its own or in combination with docosahexaonic acid as adjunctive treatment.[6] Serotonergic remedies should not be combined with each other or with conventional antidepressants because of the increased risk of serotonin syndrome. Equally, herbal antidepressants may induce mania in vulnerable patients although current evidence relies on case reports only. Finally L-tryptophan and 5-hydroxytryptophan should be avoided until the associated risk of eosinophilic myalgic syndrome is fully explained.

Remedies for psychosis

Rauwolfia (*Rauvolfia serpentina*) extracts were traditionally used before synthetic antipsychotics became widely available. Several alkaloid derivatives including reserpine were introduced in the 1950s. They block vesicular storage of monoamines so that the presence of monoamines in the cytoplasm is prolonged, enabling them to be more easily degraded by monoamine oxidases. In consequence, the amount of neurotransmitter available on depolarization of the cell membrane is reduced.[4] On the one hand, this may lead to a reduction of dopamine and the resolution of psychotic symptoms. On the other hand, less serotonin and noradrenaline will be available, which explains why drugs such as reserpine may precipitate depression. An alternative strategy is the augmentation of antipsychotic treatment with omega-3-fatty acids, but the results of clinical trials remain inconclusive and larger trials will be needed to clarify effectiveness.[6,7]

Remedies for movement disorders

Attempts have been made to treat tardive dyskinesia with antioxidants. This approach relies on the assumption that tardive dyskinesia is not only due to dopamine receptor super-sensitivity but is also related to oxidative tissue damage induced by antipsychotics. Clinical trials suggest that vitamin E may prevent the progression of tardive dyskinesia. One trial found actual improvement. However, the benefits have to be offset against taking vitamin E long-term, particularly when higher than recommended daily doses are used.[5] A far more powerful antioxidant than vitamin E is melatonin attenuating dopaminergic activity in the striatum as well as hypothalamic dopamine release.[8]

Remedies for the treatment of addiction

Only few plants have been identified as having the potential to counter addiction. Such may be ibogaine, derived from the West African shrub *Tabernanthe iboga*. It has hallucinogenic properties, and has been used to counter nicotine, cocaine, and opiate addiction. It causes dose-dependent CNS stimulation ranging from mild excitation and euphoria to visual and auditory hallucinations. The therapeutic value of ibogaine is limited since it is highly neurotoxic and can cause irreversible cerebellar damage. A synthetic derivative with similar reported effects, but without cerebellar toxicity is 18-methoxycoronaridine (18-MC).[4] Between 1990 and 2006, twelve deaths after ibogaine use were reported. More deaths may have occurred but may not have been reported due to the 'underground nature of ibogaine treatment'. Passion flower and valerian, by virtue of their GABAergic properties, may ameliorate withdrawal symptoms. Kudzu, Japanese arrowroot (*Pueraria lobata*), has traditionally been used for the treatment of alcohol hangover. The active ingredient, puerarin, counteracts the anxiogenic effects associated with alcohol withdrawal.[9]

Examples of remedies commonly used for chronic somatic conditions

Many different remedies are available for somatic conditions. Their use may be problematic in chronic conditions such as cancer or HIV where the therapeutic margin of conventional medicines is narrow.[10] For example, echinacea (*Echinacea purpurea*) is used to boost immune system. This may be detrimental where immunosuppression is desired since echinacea may potentially stimulate the growth of malignant or infectious cell lines. Patients with breast cancer may often resort to phytoestrogens such as soy (*Glycine max*), wild yam (*Dioscorea alata*), or liquorice (*Glyccyrhiza glabra*) to reverse the effects of antiestrogenic therapies such as tamoxifen. Phytoestrogens, however, can theoretically stimulate breast cancer cells and thus should be advised against in this patient group. Liquorice is a popular ingredient of many traditional Chinese medicines and may cause hypokalemia if used excessively. Evening primrose (*Oenothera biennis*) oil is a popular remedy for premenstrual syndrome and mastalgia. Like starflower it contains γ-linolenic acid and may lower the seizure threshold or reduce the efficacy of antiepileptic drugs.

Drug interactions

Determining interactions between complementary and conventional medicines can be extremely difficult. In the first instance, the clinician must be prepared to consider such a possibility and take a corresponding history. As often, association does not prove causality. Drug interactions can be distinguished into pharmacodynamic and pharmacokinetic interactions. Pharmacodynamic interactions occur when remedies act as agonists, antagonists or inverse agonists to conventional medicines. Additive toxicity, e.g. hepatotoxicity due to pyrrolizidine alkaloids, or increase of coagulability due to coumarinic constituents may also be of concern. Pharmacokinetic interactions include interactions with the cytochrome microsomal enzyme system (CYP) or membrane transporter proteins expressed through the ABC cassette genes (Table 6.2.9.2).[1,2,11,12]

The CYP system

The pharmacokinetics of most anticancer drugs is highly variable and may be genetically determined. For instance, the oxidative metabolism depends on the CYP system. The effects of CYP inducers and inhibitors are essentially differential depending on whether metabolites are more or less active (Table 6.2.9.2). If metabolites are less active than the original agent, CYP inhibitors increase whereas inducers reduce therapeutic effectiveness. Conversely, if metabolites are more active than the original agent, CYP inhibitors reduce whereas as inducers increase therapeutic effectiveness. Often such interactions have only been studied *in vitro* and it remains unclear whether they translate into tangible clinical effects.[11,12] In clinical practice, it is often possible to monitor combination of medicines more closely or to adjust the doses of conventional drugs in the required direction rather than to advise discontinuation of complementary remedies. Interactions with CYP 3A4 are of particular concern, since this enzyme metabolizes up to 60% of all clinically used drugs including HIV protease inhibitors, HIV non-nucleoside reverse transcriptase inhibitors, warfarin, ciclosporin, oral contraceptives, digoxin, theophylline, anticonvulsants, and various psychoactive drugs.[13]

ABC transporters

The ABC cassette genes represent proteins binding to ATP and use this energy to drive various molecules through cell membranes. The transport is mostly unidirectional. The ABC genes have been mainly explored for their capacity to cause multi-drug resistance in cancer chemotherapy.[12] Thus remedies exerting such effects may be of particular interests to the liaison psychiatrist. The most commonly known transporter is p-glycoprotein involved in the

Table 6.2.9.2 Examples of potential drug interactions of commonly used psychotropic remedies

Remedy	Pharmacodynamic	Pharmacokinetic
Ginkgo	Antithrombolytic agents	1A1, 1A2, 2B1/2, 2C9, **2C19***, 3A1, **3A4** inhibition
Panax ginseng	Insulin and oral hypoglycaemics, antithrombolytic agents, MAOIs (phenelzine), loop diuretics	1A1, 1A2, 1B1, 2C9, 2C19, **2D6**, 2E1 inhibition3A4 inconclusive; p-glycoprotein inhibition
Hydergine	Serotonergic antidepressants, choline-esterase inhibitors	
Vitamin E	Anticoagulants and antiplatelet drugs; prevention of nitrate tolerance possible; ↑ effect of sildenafil and related phophodiesterase-5 inhibitors possible; ↓ effect of chemotherapies relying on oxidative stress	CYP 3A11 induction
Valerian	↑ Effect of sedatives	**CYP 3A4** and p-glycoprotein inhibition
Passion flower	Anticoagulants, ↑ effect of sedatives	CYP 3A4 inhibition
Kava	↓ Effect of levodopa	Potentiation of liver toxicity of other drugs CYP 1A2, 2C9, 2C19, 2D6, 3A4 and 4A9/11 inhibition
Melatonin	Anticoagulants, ↑ effect of sedatives; ↓ effect of chemotherapies relying on oxidative stress	
St John's wort	Serotonergic antidepressants	CYP 3A4, 1A2, 2C9, **3A4**and **2E1** induction:, p-glycoprotein induction
Omega 3 fatty acids	↑ Effect of warfarin, aspirin and non- steroidal anti- inflammatory drugs	CYP 3A4 and p-glycoprotein inhibition
Rauwolfia	↑ Effect of anti- psychotics and barbiturates; ↓ effect of levodopa; severe bradycardia with digitalis glycosides; hypertension in combination with sympathomimetics	
Iboga	Cholinergic and anticholinergic drugs	
Echinacea	↓ Effect of immunosuppressants	CYP **2A1**, 2C9, and **3A4** inhibition, CYP 3A4 induction also possible depending on extract
Evening primrose oil	Other drugs reducing seizure threshold, anticoagulants	1A2, 2C9, 2C19, 2D6, 3A4 inhibition

*Bold font: *In vivo* evidence available.

transport of many psychotropic drugs through the blood brain barrier. St John's wort, valerian, and panax ginseng are remedies shown to change p-glycoprotein activity (Table 6.2.9.2).[11,12] However, whether such effects are sufficiently powerful to affect conventional treatments remains unclear.[14]

Conclusions

At present, the evidence base for the use of psychotropic complementary medicines is extremely limited. Due to the large variability of formulations it can be extremely difficult to conduct clinical trials with replicable results even if a candidate plant has been identified. Pooling results of existing trials in meta-analyses may be unhelpful if the trials are too small or heterogenous or if the analysis is not adjusted for the extract types used. Equally, systematic pharmacovigilance is difficult to implement in the absence of a regulatory framework.

Clinicians need to be aware that patients may use complementary therapies regardless of the evidence available and should inquire about such forms of self-medication. Pattern of use may vary with cultural background and health beliefs. Given the complex pattern of potential interactions, conventional health care professionals should not be afraid to discuss complementary use with their patients. For instance, complementary medicines should be considered a potential cause when the clinical presentation, the treatment result, adverse effects, or even diagnostic investigations are unusual or unexpected. Equally, patients should be encouraged to disclose information about complementary medicines to health care professionals. On the one hand, discussions need to be conducted sensitively in order to avoid alienating patients who may feel that they have not been taken seriously or have been criticized for using complementary medicines. On the other hand, uncritical encouragement of potentially harmful or inappropriate use of complementary medicines may possibly lead to litigation.(15) In most cases, remedies may not have to be discontinued if conventional treatments are closely monitored and adjusted. A constructive discussion about complementary medicines may potentially be a gateway towards enhancing compliance with conventional treatments.

Further information

Memorial Sloan-Kettering Cancer Center: Cancer Information: Integrative Medicine: www.mskcc.org. Keyword: herbs.

National Centre for Complementary Alternative Medicines / National Institute of Health: http://nccam.nih.gov.

Royal Botanic Gardens, Kew: Education: Resources: Information Sheets: www.rbgkew.org.uk/ksheets/.

Royal College of Psychiatrists: Mental Health Information: Therapies: Complementary and Alternative Medicines 1 & 2: www.rcpsych.ac.uk/mentalhealthinformation/therapies.aspx.

The Prince of Wales Foundation for Integrated Health: http://www.fih.org.uk/.

References

1. Werneke, U., Turner, T., and Priebe, S. (2006). Complementary alternative medicine in psychiatry: a review of effectiveness and safety. *The British Journal of Psychiatry*, **188**, 109–21.
2. Natural Medicines Comprehensive Database. (2007). Keyword: product search. www.naturalmedicines.com.
3. Ernst, E. (2002). Adulteration of Chinese herbal medicines with synthetic drugs: a systematic review. *Journal of Internal Medicine,* **252**, 107–13.
4. Spinella, M. (2001). *The psychopharmacology of herbal medicine*. MIT Press, Cambridge.
5. Food Standards Agency. Expert Group on Vitamins and Minerals. (2003). *Safe upper levels for vitamins and minerals*. www.foodstandards.gov.uk.
6. Freeman, M.P., Hibbelen, J.R., Wisner, L.K., *et al.* (2006). Omega-3 fatty acids: evidence base for treatment and future research in psychiatry. *The Journal of Clinical Psychiatry*, **67**, 1954–67.
7. Joy, C.B., Mumby-Croft, R., and Joy, L.A. (2003). *Polyunsaturated fatty acid supplementation for schizophrenia* (*Cochrane review*). In *The* Cochrane Library, Issue 4, John Wiley & Sons, Ltd., Chichester, UK.
8. Lohr, J.B., Kuczenski, R., and Niculescu, A.B. (2003). Oxidative mechanisms and tardive dyskinesia. *CNS Drugs*, **17**, 47–62.
9. Overstreet, D.H., Keung, W.M., Rezvani, A.H., *et al.* (2003). Herbal remedies for alcoholism: promises and possible pitfalls. *Alcoholism: Clinical and Experimental Research*, **27**, 177–85.
10. Labriola, D. and Livingston, R. (1999). Possible interactions between dietary antioxidants and chemotherapy. *Oncology* (*Williston Park*), **13**, 1003–8.
11. Sparreboom, A., Cox, M.C., Acharya, M.R., *et al.* (2004). Herbal remedies in the United States: potential adverse interactions with anticancer agents. Journal of Clinical Oncology, **22**, 2489–503.
12. Sparreboom, A., Danesi, R., Ando, Y., *et al.* (2003). Pharmacogenomics of ABC transporters and its role in cancer chemotherapy. *Drug Resistance Updates*, **6**, 71–84.
13. Committee of Safety in Medicine & Medicines Control Agency. (2000). Reminder: St John's wort (*Hypericum perforatum*) interactions. *Current Problems in Pharmacovigilance*, **26**, 6–7.
14. Morris, M.E. and Zhang, S. (2006). Flavonoid-drug interactions: effects of flavonoids on ABC transporters. *Life Sciences*, **78**, 2116–30.
15. Cohen, M.H. and Eisenberg, D.M. (2002). Potential physician malpractice liability associated with complementary and integrative medicinal therapies. *Annals of Internal Medicine*, **136**, 596–603.

6.2.10 Non-pharmacological somatic treatments

Contents

6.2.10.1 Electroconvulsive therapy

Max Fink

Introduction

Convulsive therapy (ECT or electroshock) is an effective treatment for those with severe and persistent emotional disorders. It is safe for patients of all ages, for those with debilitating systemic illnesses and during pregnancy. It relieves symptoms in a briefer time than do psychotropic drugs. To achieve remission, treatments are usually given three times a week for two to seven weeks. To sustain recovery, treatments are continued either weekly or biweekly for several months. The overall duration of the treatment course is similar to that of the psychotropic medications frequently used for the same conditions.

The treatment is severely stigmatized and its use is discouraged, even interdicted, in the belief that the electricity or the seizures irreversibly damage the brain.[1–5] Few physicians are tutored in its use and facilities are limited making ECT unavailable to many who would benefit. The ease in the use of psychotropic medications, and neither greater efficacy nor greater safety, encourages their preferential use as ECT is relegated to the 'last resort.' In countries where psychotropic medications are expensive, ECT is prescribed, but the expense for anesthetics limits its use to its unmodified form.

Despite these hurdles of stigma, expense and lack of training, its use has persisted for more than 70 years. Indeed, its use is increasing. Whole societies where it was interdicted at the end of the 20th century, as in the Netherlands, Germany, Austria, Italy, and Japan, interest and usage has increased, texts have been written or translated, and local psychiatric societies formed to encourage its use.[4–6]

Origins

At the turn of the century when malarial fevers were used to treat patients with neurosyphilis, it was deemed possible to treat one illness by developing another. Reports that patients with dementia praecox were relieved of their psychosis after suffering convulsions supported a concept of an antagonism between epilepsy and psychosis. An explanation was seen in the reports that the concentrations of brain glial cells in patients with dementia praecox were low and in those with epilepsy were high.[7] Was it possible that the root of schizophrenia lay in the paucity of glia and would their increase relieve the illness?

After testing ways to induce a grand mal seizure in animals, Ladislas Meduna, a Hungarian neuropathologist and psychiatrist, on January 24, 1934, injected camphor-in-oil into a man with the catatonic form of schizophrenia. The patient seized and recovered without incident. Following the model of malarial fever therapy, Meduna repeated the injections every three days, and after the fifth seizure, the patient, for the first time in four years, talked spontaneously and fed and cared for himself. After three additional treatments he was discharged home, returned to work, and was well when Meduna left Hungary in 1939.[7]

Chemically induced seizures, either with camphor or pentylenetetrazol (Metrazol), were rapidly adopted worldwide as the treatment for dementia praecox, but the treatments were painful and

frightening, so alternative means were sought. In 1938, the Roman psychiatrists Ugo Cerletti and Luigi Bini demonstrated the ease of administration and the efficiency of electrically induced seizures. Quickly, 'electroconvulsive therapy' (ECT, electroshock) became the commonest method of inducing seizures and is the standard induction today.[4–5,8]

For whom is ECT effective?

Established DSM diagnoses are usually cited as the indications for ECT. The diagnoses are imprecise, however, offering heterogeneous population samples. A syndromic view offers more homogeneous populations for treatment.[9–11]

Defined by DSM Classification. The DSM defined conditions for which ECT is prescribed are cited in established texts[12–15] (Table 6.2.10.1.1). The breadth of its clinical efficacy across major DSM diagnostic classes is striking, reflecting commonalities in the pathophysiology of different disorders. This experience challenges the concept that DSM classified disorders are distinct biological abnormalities, and supports the 19th century concept of a unitary psychosis.[16]

ECT is ***not*** useful for a patient with neurosis, situational maladjustment, personality disorder (character pathology), or drug dependence. It is of limited benefit for anyone with a lifelong history of mental and emotional dysfunction, unless the onset of the present illness is acute and well defined, or affective, psychotic, or catatonic features dominate the presentation[12–15] (Table 6.2.10.1.2).

Defined by syndrome. DSM-III and DSM-IV classify illnesses based on the check-off of symptoms modified by duration criteria. The DSM criteria identify heterogeneous populations that do not support useful treatment algorithms or the search for biological roots of the illnesses. Clinical syndromes describe more homogeneous populations, often substantiated by biological tests and/or a high specificity of interventions. Melancholia, psychotic depression, catatonia, delirious mania, and acute schizophrenia are syndromes that are particularly responsive to ECT (Table 6.2.10.1.3). These syndromes are not readily identified in the established classification systems. Summary descriptions are offered here; the interested reader will find more extensive descriptions in the cited literature.

(a) Depressive mood disorders

Convulsive therapy is most effective against mood disorders, depression and mania. Depressive mood disorders are dominated by sadness, hopelessness, fears, and thoughts that life is no longer worth-living. Variants are recognized, dominated by vegetative and motor abnormality (melancholia), by delusions (psychotic depression), by severe cognitive deficit (pseudodementia), or by catatonia.[14]

While all variants respond to induced seizures, some also respond to other specific treatments. Melancholic and pseudodementia patients respond to tricyclic antidepressants. Psychotic depressed patients require high doses of both antidepressant and antipsychotic medications.[11] Anticonvulsant sedative drugs, the barbiturates and the benzodiazepines, are useful in catatonic patients.[9]

(i) Melancholia

Motor signs (retardation or agitation) and vegetative symptoms of inability to sleep, feeding, and weight loss are its features. Work, sex, and family are disregarded. Thoughts of suicide are prominent.[10–11]

Table 6.2.10.1.1 DSM defined clinical diagnoses in which ECT is effective[11]

Major depressive disorder	
single episode	[296.2x]**
recurrent	[296.3x]**
Bipolar disorder	
mania	[296.4x]**
depressed	[296.5x]**
mixed type	[296.6x]**
not otherwise specified	[296.70]**
Atypical psychosis	**[298.90]**
Schizophrenia	
catatonia	[295.2x]
schizophreniform	[295.40]
schizo-affective	[295.70]
Catatonia	
Schizophrenia, catatonic type	[295.2x]
Catatonic disorder due to a medical condition	[293.89]
Malignant catatonia	[293.89]
Neuroleptic malignant syndrome	[333.92]
Delirium	
Due to a general medical condition	[293.0]
Due to substance intoxication	

* from Fink, 1999

** specifier for psychosis

Hypercortisolemia is characteristic of the syndrome.[17] Cortisol metabolism is influenced by hypothalamic, pituitary, and adrenal interactions. Melancholic patients exhibit elevated serum levels of cortisol, obtunded diurnal rhythmicity, and serum levels remain elevated despite an administered dose of dexamethasone. The abnormality is measurable by the dexamethasone suppression test (DST) or its variant, the dexamethasone/corticotrophin releasing factor test (Dex/CRH). Elevated cortisol levels normalize with treatment and become abnormal again with relapse. In the 1980s, the specificity of the DST was considered poor for the major depressions defined by DSM-III and the test was discarded. But the re-assessment of the literature and recent reports find the test as

Table 6.2.10.1.2 DSM diagnoses in which ECT is ineffective[11]

Dementia and Amnestic Disorders	[293.0, 290.xx, 294.xx]
Substance-related Disorders	[303.xx, 291,x, 304.x, 292.x]
Anxiety and Somatiform Disorders	[300.xx]
Factitious Disorders	[300.xx]
Dissociative Disorders	[300.1x, 300.6]
Sexual Dysfunctions	[302.xx, 625.8, 608.89, 607.84, 608.89, 625.8]
Sleep Disorders	[307.xx, 780.xx]
Impulse disorders	[312.3x]
Adjustment disorders	[309.xx]
Personality disorders	[301.xx]

Table 6.2.10.1.3 ECT responsive syndromes

Mood disorders
Depression
Melancholia
Psychotic (Delusional) depression
Mania
Mixed States (mania, depression)
Rapid cycling mania
Depressive phase of bipolar disorder
Delirious mania
Psychosis
Acute schizophrenia
Postpartum psychosis
Catatonia
Hypokinetic catatonia (Kahlbaum Syndrome)
Excited catatonia (delirious mania, oneiroid state)
Malignant catatonia (NMS, TSS)
Other
Delirium
Suicide risk
Status epilepticus (SE< NCSE)

* from Fink, 1999.

both sensitive and specific for melancholic depression, where it has a positive predictive value.[11,18]

After an extensive review of the literature, Taylor and Fink (2006) concluded that classifying mood disorder patients as either melancholic or non-melancholic offered more homogeneous populations with better outcomes with TCAs and ECT than did the DSM classification of major depression and bipolar disorder.[11] In their formulation, melancholia is a syndrome of depressive mood, with motor and vegetative abnormalities and with evidence of cortisol abnormality.

(ii) Delusional (psychotic) depression

Overwhelmed by feelings of helplessness, hopelessness, and worthlessness, the patient believes others are watching or talking about him, reporting voices when no one is present. He imagines that events depicted on a television or movie screen apply directly to him. This form is labelled psychotic depression and is remarkably responsive to ECT.

In 1975, Glassman and his associates at Columbia University reported that only three of 13 delusional depressed patients (23 per cent) improved when they were treated with high doses of imipramine, while 14 of 21 non-delusional patients (66 per cent) improved under the same treatment.[18] Nine of the 10 unimproved delusional patients responded well to ECT. These findings have been repeatedly verified.[11,12]

In a study of 437 depressed hospitalized patients treated with imipramine in doses of 200 to 350 mg/day for 25 days or longer, 247 (57 per cent) were evaluated as recovered and were discharged.[19] When the 190 unimproved patients were treated with bilateral ECT, 156 (72 per cent) were recovered. Most of the depressed patients who had not improved with imipramine were delusional as well as depressed.

Only a third of delusional depressed patients recover when treated with antidepressant drugs alone and half recover with antipsychotic drugs alone.[11, 19–20] Two-thirds of those treated with ECT or with high doses of both antidepressant and antipsychotic drugs regain their health.

In a two-year study of late-life depression, 47 per cent of the delusional depressed patients treated with medication relapsed earlier and more often than the nondelusional depressed (15 per cent), indicating that delusional depression is particularly resistant to medication.[21] It is, however, so amenable to ECT that it is considered a primary indication for its use.[11–14] But the condition is difficult to diagnose making inadequate treatment common. In a three-hospital research study of ECT and continuation medications, only 2 of 52 delusional depressed patients had adequate courses of medication treatment before they were referred for ECT.[22] The same failure was found in another multi-center study with only 5 of 106 patients failing adequate courses of treatment before referral to ECT.[23]

Many reviews find psychotic depression to have a more severe pathophysiology and just using the same treatments as for non-psychotic depression, even at much higher doses is not adequate.[11,12] Yet, bilateral ECT is remarkably effective. In a multi-site collaborative ECT study, of 253 patients with unipolar major depression, 77 were psychotic depressed. Their remission rate was 95 per cent compared to 83 per cent for the non-psychotic depressed, with the speed of response faster for the psychotic depressed patients.[24, 25]

(iii) Pseudodementia (reversible dementia)

Because the depressed patient ignores daily events, little of what happens to him is registered and memory is compromised. The condition is hardly distinguishable from Alzheimer's dementia. The onset is usually more rapid and severe compared to the onset of a structural dementia, and patients often report a history of prior depressive episodes.[11,14]

Because the syndrome is not well known, patients are often sent to nursing home care. An example of a 58-year old woman who developed a reversible dementia and was not adequately treated for eight years is reported. Once the diagnosis was considered, antidepressant treatment relieved the syndrome and returned the patient to a more normal family life.[14]

(iv) Catatonia

When the patient is mute, sitting rigidly in a chair or lying motionless on his bed, and unresponsive to questions and commands, he appears as in a stupor. The state is called *catatonia* or *depressive stupor*. Catatonia is seen among patients with many DSM diagnoses.[9] It is discussed in detail below.

(b) Manic mood disorders

A mood disorder dominated by grandiosity, expansiveness, feelings of increased power and energy, and excitement, can last for hours, days, weeks, or months. Even after it is relieved, it may recur or alternate or combine with episodes of depression. When the switches occur within one or a few days, the experience is labelled *rapid cycling*, a malignant form of the illness. *Bipolar disorder* is the label applied to both mania and mixed forms of the illness.[11,26]

Disturbances in eating and sleeping, thinking, memory, and movement are features of mania. The patient does not sleep, eats poorly, loses weight, and concentrates thoughts poorly. Memory is impaired, often severely; he may be so disorganized as to appear demented and delirious. Melancholia, psychosis, pseudodementia, and catatonia variations are commonly seen.

Delirious mania is a striking form of mania. A normal person suddenly becomes excited, restless, and sleeps poorly, fears that neighbors are watching him, and is easily frightened. He may hide in the house or leave it abruptly, dressed inappropriately, sometimes naked, and wander about the streets. His hallucinations are vivid, his thoughts disorganized. Confusion alternates with mutism, posturing, rigidity, and stereotyped repetitive movements. Physical exhaustion even to the point of death occurs.[11,27]

Before ECT, patients were sedated with opiates, bromides, or chloral and many died of poor care, inanition, and pneumonia. A 1994 summary of the reports of manic patients treated with ECT finds 371/562 (66 per cent) remitted or showed marked clinical improvement.[28] The introduction of chlorpromazine and other sedative drugs quickly replaced ECT for efficacy and ease of use. But when chlorpromazine and other antipsychotic drugs were used in place of ECT, the doses often carried the risks of sudden death and neuroleptic malignant syndrome, as well as tardive dyskinesia and tardive dystonia.[9]

Anticonvulsant drugs are now preferentially recommended, even though the evidence for their efficacy is poor. Many authors encourage the use of lithium for immediate relief and for prophylaxis. In 438 manic patients treated with ECT or lithium, 78 per cent of the ECT treated group showed marked improvement compared to 62 per cent of those treated with adequate doses of lithium and 56 per cent of those treated with inadequate doses.[29] The group receiving neither ECT nor lithium fared least favourably with only 37 per cent improved.

No matter the array of medications and polypharmacy for mania, ECT is an effective alternative.

(c) Catatonia

Muscular rigidity, posturing, negativism, mutism, echolalia, echopraxia, and stereotyped mannerisms, the signs of catatonia, appear suddenly and immobilize patients.[9] When the disorder is transient, it may be disregarded, but when it persists, it threatens life. Patients undergo forced feeding and develop bedsores, muscular atrophy and pulmonary embolization. Repeated bladder catheterizations induce infections.

Catatonia is recognized in patients with affective illnesses, both depression and mania, in patients with systemic disorders, and in those with toxic brain states caused by hallucinogenic drugs. For decades, the prevailing belief was that each instance of catatonia represented schizophrenia. The major classification systems in psychiatry — DSM-III and IIIR of the American Psychiatric Association and the International Classification of Diseases (ICD-IX, ICD-X) — assigned patients with catatonia to the diagnosis of schizophrenia, catatonic type. Few patients were treated with anticonvulsant sedatives or ECT, despite their known efficacy, because neither was recommended for schizophrenia. This short-sighted view was somewhat corrected in the 1994 classification system of the American Psychiatric Association (DSM-IV), which recognized catatonia as secondary to systemic illness in the class of "*Catatonic disorder due to . . . . (Indicate the General Medical Condition)* [293.89]".[30] The experience that catatonia is not limited to patients with "schizophrenia" has led to the call for a separate category in DSM-V.[9,31]

Catatonia is defined by the persistence of two or more characteristic motor signs for more than 24 h in a patient with a mental disorder.[9,31] Posturing and staring can be observed, but most signs require elicitation in the examination. The accepted motor signs and a formal examination are cited in catatonia rating scales.[9] An intravenous challenge of lorazepam or amobarbital verifies the diagnosis in more than 2/3 patients with catatonia, and a positive test response augurs well for high dose benzodiazepine therapy. When this treatment fails, ECT is effective, although the treatment schedule may require daily treatments.

Catatonia may be transitory or may persist for months or years. It appears in many guises.[9,32] Prominent examples are *malignant (pernicious) catatonia* (MC) with a high risk of death and the *neuroleptic malignant syndrome* (NMS) that follows on the administration of neuroleptic drugs.

(i) Malignant catatonia

Descriptions of patients who develop an acute febrile delirium with excitement or stupor dot the literature. They often exhibit signs of catatonia. Vegetative dysregulation is often severe and death was a frequent feature before the introduction of ECT. Descriptions by Bell (1849), Stauder (1934), and Bond (1950) highlight the lethal nature of the syndrome. In 1952, Arnold and Stepan described patients in whom ECT rapidly relieved malignant catatonia, but to avoid mortality it had to be used within the first five days.[9]

(ii) Neuroleptic malignant syndrome (NMS)

A toxic response to neuroleptic drugs evinced by fever, motor rigidity, negativism, mutism, and cardiovascular and respiratory instability is a toxic response to neuroleptic drugs. It is indistinguishable from malignant catatonia.[9,32] It is an MC variant as the diagnostic criteria and effective treatments are the same as for MC. MC occurs with almost all neuroleptics, most commonly with the high-potency agents like haloperidol, fluphenazine, and thiothixene, but also with atypical neuroleptics.

One hypothesis explains the syndrome as a consequence of an excessive reduction in the amount of brain dopamine. Those who believe this association prescribe the dopamine agonists bromocriptine or levodopa and relieve muscular rigidity by prescribing the muscle relaxant dantrolene. Neither of these treatments has proved effective and dantrolene use is associated with considerable toxicity.[33] These are best not used and patients are best treated with sedative anticonvulsants and ECT.

(iii) Toxic serotonin syndrome (TSS)

A toxic syndrome is occasionally described in association with the SSRI antidepressant drugs. TSS is similar to MC with prominent gastrointestinal symptoms. The diagnosis and treatment follows the protocol for MC.[9]

(d) Psychosis

A severe impairment of thought characterized by delusions is a feature of many psychiatric conditions, notably manic delirium, psychotic depression, post partum depression, and toxic psychosis. It is broadly defined as a psychosis and diagnosed within the major class of psychoses as schizophrenia. In this class ECT is hardly considered. But when we consider the efficacy of ECT in the psychotic variants of the mood disorders, we appreciate that ECT is an effective treatment of psychosis.[34]

Convulsive therapy was introduced for the treatment of dementia praecox and was widely and quickly adopted. Comparisons with chlorpromazine found both treatments effective in acute and severe short-term illnesses, but neither was useful in chronic states. Chlorpromazine was favoured since its cost is considerably less and

its image better. As more patients failed to respond to medications, however, a cadre of 'medication resistant' psychotic patients developed. Families asked whether anything else could be done to better the patients' lives. Friedel (1986) augmented a failed course of thiothixene therapy with ECT, returning each of nine patients to community life. The finding was replicated in the successful augmentation in 8/9 psychotic patients.[34]

Clozapine was described as a treatment for psychotic patients who had failed to respond to two different antipsychotics. As the experience with this treatment grew, clinicians were again faced with treatment failures and ECT augmentation was tried. A synergy for ECT and clozapine was described and offers an effective treatment for patients who have failed conventional antipsychotics and clozapine.[34]

It is reasonable to consider ECT in the treatment of psychosis, whether in an affective illness or in schizophrenia. For the affective illnesses, ECT is used alone. In schizophrenia, ECT is effective alone or in augmenting neuroleptics.[34]

(e) Delirium

Acutely ill psychotic patients often exhibit disturbances in consciousness and are confused. Delirium is common in toxic states, either drug induced (alcohol being the most common), or secondary to drug withdrawal, or associated with systemic illnesses. Delirium is a feature of acute manic states (e.g. delirious mania) and the confusional state described as oneirophrenia. With few resources to treat acute psychoses, ECT was applied with favourable results.[14,35] The relief of delirium by ECT is an unrecognized effect that warrants consideration as an alternative to the risks of high potency neuroleptic drugs inducing NMS (MC).

(f) Neurological syndromes

ECT is well appreciated in catatonia, but it is also useful in status epilepticus (SE), non-convulsive status epilepticus (NCSE), and Parkinsonism.

(i) Status epilepticus

SE and NCSE are emergency conditions with high mortality rates. The pathophysiology is the persistence of seizures as biochemical inhibitory mechanisms fail to terminate a seizure.[36] Despite ever larger doses of anticonvulsant medications, proceeding from lorazepam to phenytoin, phenobarbital, and general anesthesia with midazolam, propofol, or barbiturates, patients persist in SE and NCSE.

ECT is another effective intervention. During the course of electroconvulsive therapy, the seizure threshold rises, encouraging seizure termination. The first report of the relief of intractable epilepsy by ECT in 1943 has been sporadically verified.[37]

An explanation for this application is physiologically interesting. The strength of a seizure can be judged by the immediate rise in serum prolactin after a sustained epileptic seizure. Within the hour after a seizure, the level of serum prolactin indicates whether the seizure is a cerebral grand mal event or a pseudoseizure. Serum prolactin levels do not rise in SE but remain normal. This suggests that the SE seizures are partial or incomplete and that they fail to stimulate an inhibitory termination process. But even in patients in SE, ECT elicits maximal seizures, making it a reasonable alternative to general anesthesia as a treatment for intractable seizures.

(ii) Parkinsonism

In treating older depressed patients with concurrent Parkinsonism with ECT, motor and facial rigidity were also relieved. In Parkinsonism, brain dopamine levels are reduced, making dopamine agonists effective treatments. In ECT, brain and CSF levels of dopamine increase. Experiments in Parkinsonian patients without mood disorder found motor rigidity to be relieved.[38] For those patients who are not relieved by conventional treatments, periodic ECT has been helpful. Continuation treatments, like continuation pharmacotherapy, are necessary to sustain the benefit.

(g) Suicide

All psychiatric disorders carry the risk of suicide. ECT reduces this drive. The impact of medications on suicide risk is not well defined but compared to ECT, the efficacy is less favourable.[6,11] Comparisons of ECT and TCAs across different treatment eras find the frequency of suicides decreased in the ECT era. A study of the psychiatric status of 519 patients six months after discharge from hospital treatment for depression found 0.8 per cent of the ECT treated patients had made a subsequent suicide attempt compared to 4.2 per cent for those rated as receiving adequate and 7 per cent of those receiving inadequate courses of antidepressant drugs. At the 6-month follow-up no suicides were reported in 34 women treated with ECT, but two suicides occurred in the 84 patients treated with antidepressants (2.4 per cent).[39]

In a study of the expressed suicide intent (changes in Item 3 of the HAMD rating scale) in 148 patients treated with ECT, the baseline average score was 1.8. It reduced to 0.1 in 72 responders and to 0.9 in 76 non-responders. For the total sample, there was a greater decrease in the suicide item scores than in the overall HAMD scores.[40]

In another study of 444 patients referred for ECT, 131 had high expressed suicide intent scores.[6] The scores dropped to zero in 106 (80.9 per cent) with treatment, occurring in 38.2 per cent (50/131) after 3 ECT (one week), in 61.1 per cent (80/131) after 6 ECT (two weeks); and in 76.3 per cent (100/131) after nine ECT (three weeks).

ECT's effect on the death rate in the mentally ill, particularly those with mood disorders, must be a major consideration in treatment recommendations.

Principles of treatment

When to consider ECT? Psychotropic drugs and psychotherapy are the first treatments of the psychiatrically ill, with referral to ECT when these treatments fail. Since ECT is effective in medication treatment failures, would it not be wise to spare patients a prolonged illness and risks of suicide by offering ECT as the initial treatment? ECT is indeed considered the first treatment when there is a need for a rapid, definitive response, as in suicidal patients who require constant observation and restraint, in hyperactive patients who may be at risk of harm to themselves or others, in those with malignant disorders as malignant catatonia, neuroleptic malignant syndrome, or delirious mania, or in those whose lives are in jeopardy from systemic illness. It is also preferred in those patients who have had a prior illness that responded well to ECT or who have had a poor experience with medications.[11–15]

How many failed trials of medications are reasonable before ECT is considered? For some patients, especially those whose practitioners are not knowledgeable about ECT, medication trials

become interminable and ECT is considered only when the patient seeks care elsewhere. A reasonable guideline is derived from the experimental trials with clozapine, an agent with life-threatening risks.[41] To put patients at risk and yet obtain the possible benefits of clozapine, the researchers decided that patients should not be offered clozapine unless they had experienced two unsuccessful courses of neuroleptic treatment. A similar standard seems reasonable for recommending ECT. After patients have failed two different courses of medications at adequate doses and for adequate periods, ECT is to be considered.

Financial considerations affect the decision. If the patient is severely ill and has only a limited ability to pay for extended care, repeated unsuccessful medication trials are unwarranted. All practitioners should balance the cost of medication trials and the effective use of ECT.

Consideration of age. ECT is an accepted treatment for adults. For decades, the attitudes of child and adolescent psychiatrists precluded consideration of ECT for their patients except the most devastatingly ill. The acknowledged safety of ECT in adults relaxed prejudices against its use and led to more treatment trials. Once it became clear that the response of adolescents was similar to that of adults, the attitude changed and ECT is now an accepted treatment for adolescents with the same illnesses that are successfully treated in adults.[42]

ECT is probably effective in similar conditions in children, but their expression of mood and psychotic disorders is different than in adults and difficult to interpret. The published experience in the few children treated with ECT finds that conditions that respond in adults and adolescents also respond in children.

ECT is widely used in geriatric patients. Indeed, it is increasingly called on when the side effects of medications become intolerable and when medication trials fail. The safety of modern ECT is such that even the frailest and systemically ill elderly can be safely treated with ECT. We acknowledge no absolute contraindication to ECT other than the lack of skill of the clinicians.[11–15]

The treatment process

Consent. The referral of a patient to an ECT service starts the treatment process. As in surgical treatment, the patient and family members are educated as to the risks and processes of the treatment course, and a signed voluntary consent, witnessed by a family member if possible, is obtained. In response to the turmoil of the 1970s when a draft for an unwelcome war led to widespread questioning of authority, attacks on ECT as a forced involuntary treatment led the profession to suggest procedures for informed voluntary consent. These procedures are well established.[1,2,4,12]

An explanation of why the treatment is recommended, specific anticipated benefits and risks, the names of the responsible physicians, and a statement that the patient may, at any time, discontinue the treatment are elements of a valid consent.[4,13,14] Although voluntary consent is the basis for ECT in almost all Western countries, provisions for involuntary treatment for patients who may not be able to understand the severity of their illness nor the need for treatment is provided in state laws with courts authorizing treatment. In a few venues, surrogate consent by family members is accepted. Educational videotapes and books for laymen support the consent process.[4,12,43]

Procedures. Treatments are usually given in an equipped room with access to the in-patient wards. Increasingly, as more than half the treatments are given to out-patients, units are established with ready access to the community.[12,13]

Prior to treatment, systemic medical examinations usually advised for general anaesthesia are completed. These include complete blood count, electrocardiogram, and urinalysis. If systemic illness is present, the treatment is optimized. Often, an anaesthesiologist will examine the patient and the record before treatment, obtaining a separate anaesthesia consent. Although no medical examinations relative to the ECT process are required, some centers unnecessarily insist on pre-treatment brain scans and EEG for all patients.

Anaesthesia. When curare and succinylcholine were introduced to modify the convulsion, patients thought they were suffocating as respiratory muscles relaxed. Momentary amnesia was provided by a barbiturate and the combination of barbiturate-induced amnesia and succinylcholine muscle relaxation became standard procedure.[8,12] When psychiatric practice changed from an office to a hospital venue, and anaesthesiologists administered medications, misunderstanding of the role of anaesthesia ensued and the benefits of treatments were reduced by high anaesthetic doses that made effective seizures difficult. Present practice is detailed in anaesthesiology texts.[44]

Monitoring and electrode placement. To monitor the physiologic effects of induced seizures, EEG and ECG electrodes are applied. To monitor the motor seizure, a blood pressure cuff is usually applied to the calf of one leg, inflated before the administration of a motor relaxant to observe the motor seizure duration. Two stimulating electrodes are required for ECT. In the early years, the electrodes were applied to both temples, with the maximum energies passing through the intervening brain tissues, especially the centrencephalic structures of the hypothalamus and pituitary. Relocating the electrodes on one side of the head to avoid stimulating the dominant temporal lobe led to seizures with less immediate impact on cognition. 'Right unilateral ECT' (RUL-ECT) became popular until clinicians realized that the efficacy of such treatments was significantly less than through bilateral electrodes (BL-ECT).[8,12]

At one time we believed that any seizure was therapeutic, but we now know that this is not so. A seizure with EEG or motor durations under 20 sec rarely develops a full grand mal convulsion. At first, effective treatments were characterized as those with a motor seizure of at least 25 sec. But not all seizures of such length are effective. Seizures induced through unilateral electrodes at near-threshold energies (experimentally identified as 1.5 and 2.5 times the calibrated seizure threshold) are not as effective as seizures induced through bilateral electrode placements.[45] Energies for seizure inductions in unilateral ECT must be at least 6 to 8 times the calibrated seizure threshold to achieve equal efficacy; at such high energies the advantage in minimizing immediate memory effects is lost.[46] As there is a linear relationship between age and seizure threshold, the energy levels with modern devices that deliver brief pulse electrical currents for BL-ECT is estimated by the half-age formula.[47] In devices that deliver 500 mC of energy at 100 per cent, the energy level for the first induction is set at half the patient's age. The quality and duration of the EEG seizure are a guide to later induction energies. In present clinical practice, electrodes are applied to both temples (BT-ECT) or over the outer canthus of each eye in 'bifrontal' (BF-ECT) placement. While the

advantages of BF-ECT and BT-ECT are being assessed in large studies, their efficacy seems equivalent. There is little justification for the use of RUL-ECT in clinical practice.

We now rely on the ictal EEG to define an effective treatment, and modern ECT devices record either one or two channels of brain electrical activity. The typical ictal EEG presents a build-up of energies, then high-voltage spike activity mixed with high-voltage slow waves (3–6 Hz), followed by trains of lesser voltage slow waves, and an abrupt end to the electrical activity with electrical silence. Such EEG patterns, generally of 35–130 sec in duration, are associated with motor seizures that are 10–20 per cent shorter. If seizures do not show these well-defined phases, we repeat the treatment at different energy settings until a robust EEG sequence is elicited.[11–15]

A rise in the post-ictal serum prolactin is another index of seizure adequacy. Grand mal seizures release brain peptides into the CSF and blood. Serum prolactin, easily measured, rises rapidly reaching a peak at about 25–30 min, and falls to a baseline level within 2 h. The absence of a dramatic rise in serum prolactin is a sign of inadequate treatment.[48]

The ECT course. Occasionally a single treatment relieves a disorder, but such instances are so rare as to be noteworthy. The basic course is more often between 6 and 20 treatments. These are usually given three times a week at the onset and, after the symptoms show some relief, are reduced to twice or once a week. The resolution of catatonia (MC, NMS) is frequently accomplished in three to five treatments but these are best administered daily. Depressive disorders require 6–12 treatments for resolution. Manic and psychotic disorders require 20 or more treatments.

Discontinuing treatment at the point of immediate resolution of symptoms is associated with high relapse rates. Continuation treatment, often continuation ECT, is as essential a part of ECT management as it is for pharmacotherapy.[49,50]

Continuation treatments. High relapse rates are the most common complaint in ECT practice. When patients are given a short course of treatments, early relapse is common. Because ECT is complex, frightening, and expensive, patients seek the shortest course of treatment, and physicians accede by prescribing a limited number of treatments on referral or at the time the patient signs the consent.[4]

Short courses of treatments may relieve symptoms but relapse is quick.[49,50] Continuation treatment is necessary. Two recent studies guide present practice. In a 3-hospital collaborative study of depressed patients referred for ECT, remitted patients were randomly assigned to 6-month courses of medication. Relapse rates were 84 per cent for placebo, 60 per cent for nortriptyline and 39 per cent for the combination of lithium and nortriptyline under serum level control.[50] In the 4-hospital collaborative study, depressed patients treated with bitemporal ECT were randomly assigned to continuation with ECT or the same lithium and nortriptyline combination. The 6-month relapse rates were 32 per cent for continuation medication and 37 per cent for continuation ECT.[49] These rates are statistically indistinguishable in the two studies.

ECT and psychotropic drugs.[12, 51] With the exception of antipsychotics, we lack evidence of synergy between psychotropic drugs and ECT.[11,12,14] TCA, MAOI, and SSRI antidepressants are usually discontinued during an ECT course. Anticonvulsants and sedative drugs affect seizure thresholds and may interfere with efficacy. ECT augmentation of antipsychotics is seen as safe and effective.

When ECT is administered to a patient with clinically effective serum lithium levels, generally seen as 0.8–1.2 mEq/l, there is the risk of a post-seizure delirium. If lithium treatment is sustained during ECT, the dosages are reduced so that the serum lithium levels do not rise above 0.6 mEq/l on treatment days.

Systemic drugs, especially those used to treat cardiovascular disorders, may put the patient at risk for hypotension, ataxia, or exaggerated cognitive deficits, but these effects can be easily managed, so they are usually continued during ECT.

Inducing adequate seizures in patients who have been receiving benzodiazepines may be difficult. Intravenous flumazenil, the benzodiazepine antagonist, effectively minimizes the inhibiting effects of benzodiazepines. Such use is encouraged for patients with catatonia or mania who have been treated with benzodiazepines.[9]

Risks and contraindications

Bone fractures, tardive seizures, and cardiac arrhythmias were common risks of early ECT, but the routine use of muscle relaxation with succinylcholine markedly reduced them.[8,12] Headache, tongue injury, and post-seizure delirium continue to be systemic risks. Headaches respond to analgesics, delirium to benzodiazepines, and tongue injury can be prevented by the proper application of bite-blocs. The principal risks of ECT today are cognitive effects and unacceptable relapse rates.

In a post-seizure delirium, which occurs in about 10 per cent of the treatments, the patient is poorly aware of where he is and may thrash about and be confused. It is more common in the first and second treatments than in later ones. Reassurance, calm talk, and gentle handling of movements that might be harmful can usually allay such states. If the restlessness does pose risks, it can be calmed by intravenous diazepam.

Persistent amnesia is the most dreaded risk of ECT.[12–15] Patients usually forget the personal events that occurred during the illness and treatment. On treatment days, both the anesthesia and the seizure alter cognition, temporarily interfering with the memory of events. In the first decades of ECT use, adequate ventilation was not assured and untoward effects on cognition were profound and frequent. But changes in practice have reduced these effects. Ventilation with pure oxygen, changes in the type of electrical current and the amounts of energy, and selected electrode placement reduced the effects on cognition, so that within a few weeks after the course is over, the patients' performance on memory tests usually surpasses their pre-treatment abilities.

Contraindications. There are no systemic illnesses that preclude the administration of ECT when the treatment is clearly warranted. Some conditions — severe hypertension, uncontrolled cardiac arrhythmia, bleeding tendencies, recent myocardial infarction, increased intracranial pressure, and a brain or cerebrovascular lesion — call for special care. The case literature offers suggestions for the appropriate treatment with ECT of patients with these conditions.[12–15]

Mechanism of action

When convulsive therapy was introduced, its most prominent side effect was amnesia, and much debate centered on whether amnesia

was, in fact, the mechanism for improving thought and mood. Experiments with different electrode placements discouraged this explanation.[4,12]

Others focused on the physiologic effects of seizures, especially the changes in the interseizure EEG. Such changes were found to be necessary, but not sufficient, for recovery.[8,52] Interest in this hypothesis is revived by recent studies.[53]

Explanations based on neurohumours and their receptors are important in our present views of the action of drug therapies. These are also cited to explain the benefits of ECT. The experimental data fail to support these explanations.[8,12,54]

Meduna thought that the concentration of glia was a factor in illness and that seizures elicited increased gliosis and recovery.[6] Recent reports cite increased neurogenesis as an active brain response to induced seizures.[55]

My view is that the neuroendocrine system is the most likely agent for the clinical changes brought about by induced seizures.[8,56] Neuroendocrine dysregulation is prominent in patients with the mental disorders for which ECT is effective. Thyroid, adrenal, sex gland, and hypothalamic dysfunction are common in patients with disorders in mood, thought, motor activity, feeding, sleep, sex, growth, and maturation. Indeed, every aspect of body physiology and mental activity is affected by these glands, as exemplified by the action of the adrenal glands in depressive mood disorders.

In the severely depressed patients, the adrenal glands produce too much cortisol.[17, 18, 56] The high blood levels disrupt the normal diurnal rhythms of other glandular discharges, and the glands do not respond to the usual feedback mechanisms. The most prominent features of depression — failure to eat, loss of weight, inability to sleep, loss of interest in sex, inability to concentrate thoughts, and difficulties in memory — are distortions of the functions regulated by the neuroendocrine glands in a self-adjusting feedback.

Each seizure stimulates the hypothalamus to discharge its hormones, which causes the pituitary gland to discharge its products, which then affects the level of cortisol. The first effects of this cascade are transitory, but repeated seizures restore the normal interactions of the hypothalamic-pituitary-adrenal axis. Feeding and sleep become normal, followed by motor activity, mood, memory, and thought.

How does a seizure elicit such profound changes in physiology? In ECT, the currents from the stimulating electrodes on each temple pass through the central parts of the brain, stimulating both the hypothalamus to discharge its hormones and the centrencephalic structures to produce a bilateral grand mal seizure. (One of the flaws in unilateral electrode placement is that the currents have to take indirect routes to affect the pertinent areas of the brain.) The massive amounts of hypothalamic and pituitary hormones that enter the bloodstream during ECT are measurable within a few minutes. They circulate throughout the body, affecting all the body's cells — a compelling and welcome sign of recovery.

After some courses of ECT, the return to normal endocrine function persists. At other times, the glands revert to their abnormal activities and the mental disorder becomes evident again. In these cases, repeated stimulation of the hypothalamus and the pituitary by continuation ECT restore and sustain normal glandular functions and support a normal mental state.

Suggested replacements for ECT

Although Meduna's experiments and numerous studies of ECT and sham ECT support the seizure as evidence of the brain changes essential to a therapeutic benefit, the introduction of electricity focussed attention away from the seizure and onto electricity as the medium for the treatment's efficacy. This interest is not new. Soon after Galvani and Volta demonstrated that electric currents could stimulate nerves and muscles, medical applications were enthusiastically sought. The first electrical experiments in the mentally ill are ascribed to Gale in New York State in 1802 and Aldini in Italy in 1803.[3,16,54] Electrical experiments were publicly demonstrated by Franklin, Mesmer, and Marat in the first years of the 19th century. Little benefit was recorded and most efforts are best considered quaint explorations.[3] At the time of World War I, faradization was a treatment for hysteria and applied in the military.[57]

During the second half of the 20th century, many techniques have been suggested as replacements for ECT, the latest being transcranial magnetic stimulation (rTMS), vagus nerve stimulation (VNS), and deep brain stimulation (DBS).[58] In rTMS rapidly alternating magnetic fields are delivered to stimulate the brain. At very high intensities, a seizure may be induced and some experiments have been undertaken to compare the seizure induced by magnetic currents with those induced by electric currents. The technique, called magnetic seizure therapy (MST), is reported to have a mild antidepressant effect.[58]

In VNS an electrical stimulator is implanted in the chest wall and electrodes are threaded through the neck to the left vagus nerve. The stimulator is similar to that used to reduce seizures in patients with severe epilepsy. The side effects of hoarseness, nausea, and vomiting are common. In DBS, the stimulating electrodes are placed in the brain, a technique occasionally used in severe Parkinsonism. We lack sufficient evidence for the efficacy of rTMS, VNS and DBS in psychiatric disorders to warrant their routine clinical use.

Device manufacturers who seek a market for their products encourage the technologies. The bad image and the stigma of ECT make its replacement the basis for exploration. At the time of this writing (Spring, 2007), no evidence has been published that any of these techniques have persistent therapeutic effects, and none are replacements for ECT.

The future in ECT

Induced seizures effectively allay severe psychiatric disorders. The treatment's stigma, however, inhibits its use and research into its mechanism. When neuroscientists recognize the unique nature of the seizure — a phenomenon that is ubiquitous in animal life — and seek to understand its biology, they will then seek ways to replace the gross process of induced seizures by more acceptable interventions. Understanding the mechanism will clarify the aetiology of psychiatric disorders. ECT will be replaced when we understand its mechanism better; for the present, continued usage is assured since no alternative intervention with its efficacy and safety is in our *materia medica*.

Further information

Abrams, R. (2001). *Electroconvulsive Therapy* (4th edn.). Oxford University Press, New York

Fink, M. (1999). *Electroshock: Restoring the Mind*. Oxford University Press, New York.

Taylor, M.A. Fink, M. (2006). *Melancholia: Diagnosis, Pathophysiology and Treatment of Depressive Illness*. Cambridge University Press, Cambridge UK.

American Psychiatric Association. (2001). *Electroconvulsive Therapy: Recommendations for Treatment, Training and Privileging*. Washington DC.

Scott, A.I.F. (Ed.) (2004). *The ECT Handbook*, (2nd edn.). Royal College of Psychiatrists, London.

Ottosson, J-,O, Fink, M. (2004). *Ethics in Electroconvulsive Therapy. Brunner-Routledge,* New York.

Shorter, E. and Healy, D. (2007). H*istory of the Shock Therapies*. Rutgers University Press, New Brunswick, NJ.

References

1. Fink, M. (1991). Impact of the antipsychiatry movement on the revival of electroconvulsive therapy in the United States. *Psychiatric Clinics of North America*, **14**(4), 793–801.
2. Fink, M. (1997). Prejudice against ECT: competition with psychological philosophies as a contribution to its stigma. *Convulsive Therapy*, **13**(4), 253–65; discussion 66–8.
3. Kneeland, T.W., and Warren, C.A.B. (2002). *Pushbutton psychiatry: a history of electroshock in America*. Westport, Conn.: Praeger.
4. Ottosson, J.-O., Fink, M. (2004). *Ethics in electroconvulsive therapy.* Brunner–Routledge, New York.
5. Shorter, E., and Healy, D. (2007). *Shock Therapy: The History of Electroconvulsive Treatment in Mental Illness.* Rutgers UP; New Brunswick, in press.
6. Kellner, C.H., Fink, M., Knapp, R., *et al.* (2005). Relief of expressed suicidal intent by ECT: a consortium for research in ECT study. *American Journal of Psychiatry*, **162**(5), 977–82.
7. Meduna, L. (1985). Autobiography. *Convulsive Therapy*, **1,** 43–57; 121–38.
8. Fink, M. (1979). *Convulsive therapy: theory and practice.* Raven Press, New York.
9. Fink, M., and Taylor, M.A. (2003). *Catatonia: a clinician's guide to diagnosis and treatment.* Cambridge University Press, Cambridge, New York.
10. Parker, G., and Hadzi–Pavlovic, D. (1996). *Melancholia: a disorder of movement and mood: a phenomenological and neurobiological review.* Cambridge Unniversity Press, Cambridge, New York, USA.
11. Taylor, M.A., Fink, M. (2006). *Melancholia: the diagnosis, pathophysiology, and treatment of depressive illness.* Cambridge; New York: Cambridge University Press; 2006.
12. Abrams, R. (2002). *Electroconvulsive therapy. (*4th ed.) Oxford Unversity Press, Oxford, New York.
13. American Psychiatric Association. (2001). *Committee on Electroconvulsive Therapy.* Weiner, R.D. The practice of electroconvulsive therapy: recommendations for treatment, training, and privileging: a task force report of the American Psychiatric Association. (2nd edn.) American Psychiatric Association, Washington, DC.
14. Fink, M. (1999). *Electroshock: restoring the mind.* Oxford University Press, New York.
15. Scott, Ae. (2004). *The ECT Handbook*. (2nd edn.) London: Royal College of Psychiatrists.
16. Shorter, E. (1997). *A history of psychiatry: from the era of the asylum to the age of Prozac.* John Wiley & Sons, New York.
17. Carroll, B.J., Curtis, G.C., Mendels, J., *et al.* (1976). Neuroendocrine regulation in depression. II. Discrimination of depressed from nondepressed patients. *Archives of General Psychiatry*, **33**(9), 1051–8.
18. Fink, M. (2005). Should the dexamethasone suppression test be resurrected? *Acta Psychiatrica Scandinavica*, **112**(4), 245–9.
19. Avery, D., and Lubrano, A. (1979). Depression treated with imipramine and ECT: the DeCarolis study reconsidered. *American Journal of Psychiatry*, **136**(4B), 559–62.
20. Kroessler, D. (1985). Relative Efficacy Rates for Therapies of Delusional Depression. *Convulsive Therapy*, **1**(3), 173–82.
21. Flint, A.J., and Rifat, S.L. (1998). Two–year outcome of psychotic depression in late life. *American Journal of Psychiatry*, **155**(2), 178–83.
22. Mulsant, B.H., Haskett, R.F., Prudic, J., *et al.* (1997). Low use of neuroleptic drugs in the treatment of psychotic major depression. *American Journal of Psychiatry*, **154**(4), 559–61.
23. Rasmussen, K.G., Mueller, M., Kellner, C.H., *et al.* (2006). Patterns of psychotropic medication use among patients with severe depression referred for electroconvulsive therapy: data from the Consortium for Research on Electroconvulsive Therapy. *Journal of ECT*, **22**(2), 116–23.
24. Husain, M.M., Rush, A.J., Fink, M., *et al.* (2004). Speed of response and remission in major depressive disorder with acute electroconvulsive therapy (ECT): a Consortium for Research in ECT (CORE) report. *Journal of Clinical Psychiatry*, **65**(4), 485–91.
25. Petrides, G., Fink, M., Husain, M.M., *et al.* (2001). ECT remission rates in psychotic versus nonpsychotic depressed patients: a report from CORE. *Journal of ECT*, **17**(4), 244–53.
26. Goodwin, F.K., and Jamison, K.R. (1990). *Manic–depressive illness*. Oxford University Press, New York.
27. Fink, M. (1999). Delirious mania. *Bipolar Disorders*, **1**(1), 54–60.
28. Mukherjee, S., Sackeim, H.A., Schnur, D.B., *et al.* (1994). Electroconvulsive therapy of acute manic episodes: a review of 50 years'experience. *American Journal of Psychiatry*, **151**(2), 169–76.
29. Black, D.W., Winokur, G., Nasrallah, A., *et al.* (1987). Treatment of mania: a naturalistic study of electroconvulsive therapy versus lithium in 438 patients. *Journal of Clinical Psychiatry*, **48**(4), 132–9.
30. American Psychiatric Association. (1994). *Task Force on DSM–IV. Diagnostic and statistical manual of mental disorders: DSM–IV.* (4th edn.) Washington, DC: American Psychiatric Association.
31. Taylor, M.A., and Fink, M. (2003). Catatonia in psychiatric classification: a home of its own. *American Journal of Psychiatry*, **160**(7), 1233–41.
32. Fink, M., and Taylor, M.A. (2001). The many varieties of catatonia. *European Archives of Psychiatry and Clinical Neuroscience.*, **251** (Suppl 1), I8–13.
33. Caroff, S.N. (2004). *Catatonia: from psychopathology to neurobiology.* (1st ed.) American Psychiatric Pub, Washington, DC.
34. Fink, M., and Sackeim, H.A. (1996). Convulsive therapy in schizophrenia? *Schizophrenia Bulletin*, **22**(1), 27–39.
35. Fink, M. (2000).The interaction of delirium and seizures. *Seminars in Clinical Neuropsychiatry*, **5**(2):93–7.
36. Lowenstein, D.H., and Alldredge, B.K. (1998). Status epilepticus. *New England Journal of Medicine.*, **338**(14), 970–6.
37. Fink, M., Kellner, C.H. and Sackeim, H.A. (1999). Intractable seizures, status epilepticus, and ECT. *Journal of ECT*, **15**(4), 282–4.
38. Fall, P.A., Ekman, R., Granerus, A.K., *et al.* (1995). ECT in Parkinson's disease. Changes in motor symptoms, monoamine metabolites and neuropeptides. *Journal of Neural Transmission. Parkinson's Disease and Dementia Section*, **10**(2–3), 129–40.
39. Avery, D., and Winokur, G. (1978). Suicide, attempted suicide, and relapse rates in depression. *Archives of General Psychiatry*, **35**(6), 749–53.
40. Prudic, J., and Sackeim, H.A. (1999). Electroconvulsive therapy and suicide risk. *Journal of Clinical Psychiatry*, **60** Suppl 2, 104–10; discussion 11–6.
41. Kane, J., Honigfeld, G., Singer, J., *et al.* (1988). Clozapine for the treatment–resistant schizophrenic. A double–blind comparison with chlorpromazine. *Archives of General Psychiatry*, **45**(9), 789–96.
42. Rey, J.M., and Walter, G. (1997). Half a century of ECT use in young people. *American Journal of Psychiatry*, **154**(5), 595–602.

43. Fink, M. (1986). Informed ECT for Patients and Families. Lake Bluff: Somatics, Inc.
44. Folk, J.W., Kellner, C.H., Beale, M.D., *et al.* (2000). Anesthesia for electroconvulsive therapy: a review. *Journal of ECT*, **16**(2), 157–70.
45. Abrams, R. (2002). Stimulus titration and ECT dosing. *Journal of ECT*, **18**(1), 3–9; discussion 14–5.
46. McCall, W.V., Reboussin, D.M., Weiner, R.D., *et al.* (2000). Titrated moderately suprathreshold vs fixed high–dose right unilateral electroconvulsive therapy: acute antidepressant and cognitive effects. *Archives of General Psychiatry*, **57**(5), 438–44.
47. Petrides, G., and Fink, M. (1996). The 'half–age' stimulation strategy for ECT dosing. *Archives of General Psychiatry*, **12**(3), 138–46.
48. Abrams, R., and Swartz, C. (1990). *The Technique of ECT*. Lake Bluff: Somatics, Inc.
49. Kellner, C.H., Knapp, R.G., Petrides, G., *et al.* (2006). Continuational ectroconvulsive therapy vs pharmacotherapy for relapse prevention in major depression: a multisite study from the Consortium for Research in Electroconvulsive Therapy (CORE). *Archives of General Psychiatry*, **63**(12), 1337–44.
50. Sackeim, H.A., Haskett, R.F., Mulsant, B.H., *et al.* (2001). Continuation pharmacotherapy in the prevention of relapse following electroconvulsive therapy: a randomized controlled trial. *Journal of the American Medical Association*, **285**(10), 1299–307.
51. Fink, M., and Kellner, C.H. (1993). ECT and Drugs: Concurrent Administration. *Convulsive Therapy*, **9**(4), 237–40.
52. Fink, M., and Kahn, R.L. (1957). Relation of electroencephalographic delta activity to behavioral response in electroshock; quantitative serial studies. *A. M. A. Archives of Neurology and Psychiatry*, **78**(5), 516–25.
53. Sackeim, H.A., Luber, B., Katzman, G.P., *et al.* (1996). The effects of electroconvulsive therapy on quantitative electroencephalograms. Relationship to clinical outcome. *Archives of General Psychiatry*, **53**(9), 814–24.
54. Shorter, E., and Healy, D. (2007). *History of the Shock Therapies.* Rutgers University Press, New Brunswick.
55. Fink, M. (2004). Induced seizures as psychiatric therapy: Ladislas Meduna's contributions in modern neuroscience. *Archives of General Psychiatry*, **20**(3), 133–6.
56. Fink, M. (2000). Electroschock revisited. *American Scientist*, **88**(2), 162–7.
57. Eissler, K.R. (1986). *Freud as an expert witness: the discussion of war neuroses between Freud and Wagner-Jauregg.* International Universities Press, New York.
58. Lisanby, S.H. (2004). *Brain stimulation in psychiatric treatment.* American Psychiatric Pub, Washington, DC.

6.2.10.2 Phototherapy

Philip J. Cowen

Introduction

Phototherapy or artificial bright-light treatment, has been used in the management of a number of medical disorders including psoriasis and hyperbilirubinaemia of the newborn. From the point of view of psychiatric treatment, the notion that light might help people with certain psychological symptoms has an ancient lineage. For example, Wehr and Rosenthal[1] cite Aretaeus who suggested in the second century AD that 'lethargics are to be laid in the light and exposed to the rays of the sun (for the disease is gloom)'. In 1898, a ship's physician named Frederick Cook recorded that the 'languor' which affected members of an Antarctic expedition during the winter darkness could be relieved with bright artificial light.[1]

The first systematic study of phototherapy as a psychiatric treatment was carried out in 1984 by Rosenthal *et al.*[2] who used bright artificial light to treat patients with the newly identified syndrome of seasonal affective disorder. Seasonal affective disorder is a recurrent mood disorder in which patients experience regular episodes of depression in autumn and winter with remission in spring and summer. Since then phototherapy has become the mainstay of the treatment of seasonal affective disorder, particularly in patients with atypical depressive features such as hyperphagia and hypersomnia. Phototherapy has also been used as an investigational treatment in other psychiatric disorders but the evidence for its efficacy in these conditions less established.

Mechanism of action

Light and seasonal and circadian rhythms

Animals and humans show circadian and seasonal rhythms in aspects of their physiology and behaviour that are influenced by environmental cues or *zeitgebers*. The light–dark cycle is believed to be one of the most important *zeitgebers* regulating circadian and seasonal rhythmicity in mammals. Mammalian circadian rhythms are driven by an 'oscillator' in the suprachiasmatic nucleus of the hypothalamus. Environmental light influences the activity of this nucleus via a neuronal pathway which runs from the retina to the hypothalamus. Thus appropriately timed bright light is able to advance or delay endogenous circadian rhythms.[3]

Lewy *et al.*[4] suggested that in patients with seasonal affective disorder the delayed onset of dawn in the autumn causes endogenous circadian rhythms to become phase-delayed with respect to clock time and the sleep–wake cycle. Bright-light treatment is able to correct this abnormality by phase advancing circadian rhythms, thereby re-synchronizing them with the sleep–wake cycle. This proposal is supported by the fact that controlled trials show that in most patients morning phototherapy is more effective than evening phototherapy.[5] While this hypothesis gives a good account of how bright-light treatment might ameliorate the symptoms of seasonal affective disorder, its possible efficacy in other conditions such as non-seasonal depression is difficult to explain by this mechanism.

Light treatment and monoamines

It is possible that bright-light treatment, through its interaction with the hypothalamus, could alter the circadian activity of the monoamine neurotransmitters involved in mood regulation. For example, some studies have shown that the antidepressant effects of phototherapy can be reversed by treatments that diminish both catecholamine and serotonin neurotransmission.[6] This has been taken as evidence that the antidepressant effects of bright light are mediated via activation of serotonin and catecholamine pathways. An alternative explanation is that in the absence of concomitant drug treatment, recovered depressed patients are vulnerable to depletion of these neurotransmitters in any case. However, effects of bright light on monoamines could account for the therapeutic effects of light in mood disorders other than winter depression.

Forms of phototherapy

The most common form of phototherapy uses a light box, which contains fluorescent tubes mounted behind a translucent plastic-diffusing screen. Depending on the fluorescent tubes employed, the light emitted is either full spectrum, which contains a little ultraviolet light, or cool white light which has no ultraviolet. The light box usually rests on a table or desk at about the eye level of a seated subject. The output of different light boxes varies but is usually between 2500 and 10 000 lux. Light sources producing 10 000 lux are more expensive but allow a reduced duration of exposure (30 min compared with 120 min) to secure a therapeutic effect.[7]

Phototherapy has also been administered using head-mounted units or light visors. These instruments are attached to the head and project light into the eyes allowing subjects to remain mobile while receiving treatment. While light visors are more convenient to use than light boxes, results from placebo-controlled trials have not been encouraging.[7]

Another form of light therapy involves the use of dawn-simulating alarm clocks. These clocks are programmed to simulate the illumination that would be experienced out of doors during sunrise on a spring day.[8] In practice, the clocks begin a gradual illumination of the bedroom about 2 h before normal wake time, increasing to a maximum of about 250 lux at the point of waking. Overall the effects of dawn-simulation in the treatment of winter depression seem equivalent to those of bright-light treatment[9] and patients often find dawn-simulating clocks more convenient (although a partner sleeping in the same room may not).

Adverse effects

Generally phototherapy is well tolerated although mild side effects occur in up to 45 per cent of patients early in treatment. These include headache, eye strain, blurred vision, eye irritation, and increased tension. Insomnia can occur particularly with late-evening treatment. Rare adverse events that have been reported include manic mood swings and suicide attempts, the latter putatively through light-inducing alerting and energizing effects prior to mood improvement. Whether these rare events are actually adverse reactions to the light is uncertain. There is no evidence that phototherapy employed in recommended treatment schedules causes ocular or retinal damage.

Indications and contraindications to light treatment

Seasonal affective disorder

The best established indication for light treatment is seasonal affective disorder where patients experience autumn and winter depressions. Clinical predictors of a response to light treatment include the following:

- Increased sleep
- Increased appetite and winter weight gain
- Carbohydrate craving
- Afternoon slump in energy
- Complete remission of symptoms in the summer

Several controlled trials have assessed the efficacy of bright-light treatment in the treatment of winter depression. In a meta-analysis of nine randomized studies, Golden *et al.*[9] found a significant benefit of bright light over dim light control with an effect size of 0.84 (95 per cent confidence interval, 0.6–1.08). A similar benefit was apparent for six studies of dawn simulation which had a mean effect size of 0.73 (95 per cent confidence interval, 0.37–1.08). While these data are compelling it needs to be remembered that it is often difficult to arrange a placebo treatment that will match the therapeutic expectation of bright light or dawn simulation.

Other mood disorders

Patients with more typical melancholic symptoms (e.g. weight loss and insomnia) do less well with bright-light treatment, even when the disorder is seasonal in nature. However, bright light has also been used in the treatment of non-seasonal depression both as a sole treatment as an adjunct to more conventional therapy. The evidence for the efficacy of bright light for this indication is less established but a Cochrane review[10] suggested that morning light treatment was significantly better than control treatment when applied as an adjunct to drug treatment or sleep deprivation. Most of these studies were of short-term duration and there are suggestions that the added benefit of light therapy does not persist when treatment stops.[11] In these studies, hypomania was more common in light-treated subjects. Phototherapy may also be of benefit in other conditions characterized by depressed mood and overeating (e.g. premenstrual dysphoria and bulimia nervosa). The literature contains reports of a number of controlled trials in such disorders where light treatment has improved depression ratings. However, the difficulty of distinguishing the specific and placebo effects of bright-light treatment relative to dim light control makes the current data difficult to interpret.

Circadian rhythm disorders

Because bright light is an effective *zeitgeber* for circadian rhythms it may also have a useful place in the treatment of disorders characterized by circadian rhythm disturbances. Such disorders encompass a range of conditions including phase-delayed or phase-advanced sleep disorder, jet lag, and problems related to shift work. In addition, disturbances of the sleep–wake cycle are common in older people with cognitive impairment. There are several reports of the utility of light treatment in these conditions; however, there is a paucity of randomized trial data.[7]

Contraindications

There are no absolute contraindications to phototherapy, except the obvious caveat that since the therapeutic effect depends on retinal activation, subjects must have sufficient visual function to allow this to occur. Otherwise it would seem prudent to avoid phototherapy in patients with pronounced and untreated agitation because this symptomatology could be worsened. In addition, evening phototherapy may worsen insomnia.

A substantial minority of patients with seasonal affective disorder meet criteria for bipolar II disorder, raising the concern that phototherapy may trigger hypomania in such individuals. Particular caution might be needed in patients with a bipolar I syndrome. Some regimes of phototherapy might lead to a degree of sleep deprivation which could also destabilize mood in bipolar patients.

Interactions

One of the advantages of phototherapy in seasonal affective disorders is that the use of antidepressant drugs may be avoided. Despite this, many patients with winter depression use phototherapy concomitantly with antidepressant medication without an obvious potentiation of adverse effects. However, a case report described apparent serotonin toxicity where phototherapy was combined with selective serotonin re-uptake inhibitors.[7]

Like bright light, the pineal hormone, melatonin, also has the ability to shift the timing of circadian rhythms[2] and theoretically melatonin taken at an inappropriate time of day could offset the antidepressant effect of light. It is also possible that bright-light treatment could exacerbate the ability of some drugs (e.g. chlorpromazine, St John's Wort) to cause skin photosensitivity reactions.[7]

Effects of withdrawal of phototherapy

If a patient with seasonal affective disorder responds to light treatment, withdrawal of treatment during the period of seasonal vulnerability leads to a return of symptomatology within a few days. It may be possible, however, to lessen the daily duration of treatment particularly towards the end of winter without inducing relapse. Otherwise cessation of light treatment does not seem to cause a specific withdrawal syndrome.

Administration of phototherapy

Since the best established indication for phototherapy in psychiatry is seasonal affective disorder, the following account will describe the use of bright-light treatment in winter depression. One of the major practical difficulties in phototherapy is the time needed to administer treatment. For this reason a 10 000 lux light box may be preferred because the daily duration of treatment can be reduced to 30 min. It seems likely that cool-white light and full-spectrum light have equivalent clinical efficacy, but because cool-white light is free of ultraviolet light it is theoretically safer and should be preferred.

The balance of evidence suggests that bright-light treatment of winter depression is most effective when administered in the early morning.[4] However, treatment given later in the day may be effective for some patients. In an initial trial, therefore, it is best to recommend early morning treatment but to advise the patient that the timing of therapy can eventually represent a balance of therapeutic efficacy and practical convenience. Treatment in the late evening should be avoided because of the possibility of sleep disruption.

Early-morning phototherapy should start a few minutes after waking. Subjects should allow themselves a 30 min duration of treatment with a 10 000 lux light source. They should seat themselves about 30 to 40 cm away from the light box screen. They should not gaze at the screen directly but face it an angle of 45° and glance across it once or twice each minute.

The antidepressant effect of light treatment usually appears in a few days but in controlled trials up to 3 weeks can be needed before the therapeutic effects of bright light exceed those of placebo treatment. If no benefit is noted after the third week of therapy, light treatment should probably be abandoned. As noted above mild side effects are common in the early stages of treatment but usually settle without specific intervention. If they are persistent and troublesome the patient can sit a little further away from the light source or reduce the duration of exposure. Exposure should also be reduced or stopped if elevated mood occurs.

Once a therapeutic response has occurred it is usually necessary to continue phototherapy up to the usual time of natural remission, otherwise relapse will occur. It may be possible, however, to lower the daily duration of treatment. Phototherapy can also be started in advance of the anticipated episode of depression as this may have a preventative effect; however, the evidence for this is limited and doubts have been expressed.[7,12]

Further information

Centre for Environmental Therapeutics. http://www.cet.org.

Seasonal Affective Disorder Association. http://www.sada.org.uk.

Eagles, J.M. (2003). Reading about seasonal affective disorder. *The British Journal of Psychiatry*, **182**, 174–6.

References

1. Wehr, T.A. and Rosenthal, N.E. (1989). Seasonality and affective illness. *The American Journal of Psychiatry*, **146**, 829–39.
2. Rosenthal, N.E., Sack, D.A., Gillin, C., *et al.* (1984). Seasonal affective disorder: a description of the syndrome and preliminary findings with light therapy. *Archives of General Psychiatry*, **41**, 72–80.
3. Arendt, J. and Broadway, J. (1987). Light and melatonin as *zeitgebers* in man. *Chronobiology International*, **4**, 273–82.
4. Lewy, A.J., Sack, R.L., Miller, S.L., *et al.* (1987). Antidepressant and phase-shifting effects of light. *Science*, **206**, 710–13.
5. Terman, J.S., Terman, M., Lo, E.S., *et al.* (2001). Circadian time of morning light administration and therapeutic response in winter depression. *Archives of General Psychiatry*, **58**, 69–75.
6. Neumeister, A., Turner, E.H., Matthews, J.R., *et al.* (1998). Effects of tryptophan depletion versus catecholamine depletion in patients with seasonal affective disorder in remission with light therapy. *Archives of General Psychiatry*, **55**, 524–30.
7. Eagles, J.M. (2004). Light therapy and the management of winter depression. *Advances in Psychiatric Treatment*, **10**, 233–40.
8. Terman, M. and Terman, J.S. (2006). Controlled trial of naturalistic dawn simulation and negative air ionization for seasonal affective disorder. *The American Journal of Psychiatry*, **163**, 2126–33.
9. Golden, R.N., Gaynes, B.N., Ekstrom, R.D., *et al.* (2005). The efficacy of light therapy in the treatment of mood disorders: a review and meta-analysis of the evidence. *The American Journal of Psychiatry*, **162**, 656–62.
10. Tuunainen, A., Kripke, D.F., and Endo, T. (2007). Light therapy for non-seasonal depression. http://www.cochrane.org.reviews/en/ab004050.html.
11. Martiny, K., Lunde, M., Unden, M., *et al.* (2006). The lack of sustained effect of bright light in non-seasonal major depression. *Psychological Medicine*, **36**, 1247–52.
12. Partonen, T. and Lonnqvist, J. (1996). Prevention of winter seasonal affective disorder by bright-light treatment. *Psychological Medicine*, **26**, 1075–80.

6.2.10.3 Transcranial magnetic stimulation

Declan McLoughlin and Andrew Mogg

Introduction

Transcranial magnetic stimulation (TMS) is a means of non-invasively stimulating the cerebral cortex using a hand-held coil applied to the scalp. In recent years TMS has been increasingly used to target neuronal circuitry implicated in neuropsychiatric disorders.

A key milestone in the development of TMS occurred in 1831 when Michael Faraday discovered the phenomenon of electromagnetic induction whereby a time-varying magnetic field can induce electrical currents through a conductor lying in proximity to the field. The French biophysicist D'Arsonval in 1896 induced phosphenes, vertigo, and syncope in human subjects by placing an induction coil around their heads. In the late 1950s, Kolin stimulated peripheral nerves (the frog sciatic nerve) with a magnetic field and a few years later the same technique was used in human subjects, inducing muscle twitching by applying a pulsed magnetic field over the ulnar, peroneal, and sciatic nerves.

In the mid-1980s, Barker and colleagues in Sheffield developed a magnetic stimulator to directly stimulate the human motor cortex.[1] They applied a circular coil, through which a large (4000 A), brief (110 μs) current was passed to the scalp. The resulting pulsed magnetic field was used to stimulate the motor cortex, evoking movements in the contralateral limbs and is known as transcranial magnetic stimulation. This ability to non-invasively stimulate the motor cortex with a magnetic field soon replaced high-voltage transcutaneous electrical stimulation for assessing central motor conduction times and mapping corticospinal pathways in a variety of neurological conditions. In the late 1980s machines capable of delivering multiple TMS pulses were developed. Repetitive TMS (rTMS), unlike single pulses of TMS can produce effects that last after the period of stimulation. For example, it has been shown that rapid rTMS (at frequencies of 5 Hz and greater) enhances motor excitability whereas slow rTMS (at 1 Hz or less) transiently depresses excitability.[2] The underlying principle of rTMS treatment is that the normal balance of excitatory and inhibitory processes within certain neuronal pathways may be disrupted in psychiatric conditions such as depression. Stimulating the brain using rTMS provides a means of increasing and decreasing excitation and inhibition in these pathways, having a neuromodulatory effect and allowing a focal targeting of specific neuronal circuitry.

Mechanism of action

The underlying mechanisms of the effects of rTMS remain poorly understood. This is in part because, as with attempting to understand the mode of action of psychotropic medication, it is difficult to establish links between cellular and physiological changes and alterations in emotion, thinking, and behaviour. Techniques used to try to better understand the molecular and physiological effects of rTMS have included neuroimaging and animal studies.

Neuroimaging has demonstrated that rTMS may exert effects on the brain at a considerable distance from the site of stimulation. For example, serial positron emission tomography scanning has been used to measure regional cerebral blood flow in medication-free patients with major depression before and after courses of fast and slow rTMS administered over the left prefrontal cortex.[3] It has been demonstrated that fast rTMS causes increases in regional cerebral blood flow in bilateral frontal, limbic, and paralimbic areas whereas slow rTMS caused decreases in blood flow in the right prefrontal cortex, left medial temporal cortex, and left basal ganglia and amygdala.

It has been suggested that rTMS of the left prefrontal cortex may modulate brain function by an effect on dopamine release. Elevated extracellular dopamine concentrations in the dorsal hippocampus have been demonstrated in the brains of rats who received rTMS. However, one of the problems with using animal models of rTMS is that currently small rTMS coils are not available and it is therefore impossible to focally stimulate one particular area of the small rodent brain. In humans it has been shown that rTMS to the dorsolateral prefrontal cortex can induce the release of dopamine in the ipsilateral caudate nucleus.[4]

Side-effects

Being non-invasive and not requiring a general anaesthetic, rTMS is considered to be a relatively safe treatment and few side-effects have been reported. The most significant potential side-effect is the risk of unintended seizure induction. There have been six reports to date of seizure induction in healthy volunteers. In half of these, very high stimulation intensities and frequencies were used. There have only been three reports of seizures in patients receiving rTMS and one of these patients had a pre-existing diagnosis of temporal lobe epilepsy. Researchers generally follow safety guidelines that exclude high-risk patients, (e.g. those with a stroke, brain tumour or pre-existing epilepsy) from receiving rTMS. These guidelines also suggest limits to the intensity, frequency, and stimulus duration of the rTMS used.[5]

The most common side-effect of rTMS is headache or facial discomfort that is the result of direct stimulation of muscle and nerves in proximity to the coil. Approximately 10–30 per cent of subjects experience these symptoms, which are generally short-lived and well-tolerated.

Technique

rTMS equipment comprises a stimulator unit, booster modules, a laptop computer, and a figure-of-eight coil. The stimulator unit contains the charging circuitry, energy storage capacitors, control electronics and discharge, and safety circuitry. It is connected to booster modules, which charge the high-voltage capacitors, enabling trains of high-intensity magnetic stimulation to be produced. The stimulating coil consists of tightly wound copper wire in a figure-of-eight through which a rapidly alternating electric current passes to produce a pulsed magnetic field. Various stimulation parameters including train duration, frequency of stimulation, stimulus intensity, and length of inter-train interval can be altered using computer software.

rTMS treatment is delivered via the figure-of-eight coil applied to the scalp surface. Typically, prior to treatment, TMS will be used

to map the motor area of the right abductor pollicis brevis (APB), and measure its motor threshold. The stimulus intensity delivered during treatment is then calculated in relation to this motor threshold. The main method of localizing the stimulation site has been to use a fixed point in anatomical relation to a specific motor area, for example the dorsolateral prefrontal cortex has generally been defined as the point 5 cm anterior to the APB motor area in the parasaggital plane. More recently some studies have used magnetic resonance imaging to more accurately delineate the area to be stimulated.

rTMS and depression

Transcranial magnetic stimulation was first postulated to have potential applications in psychiatry by Bickford and colleagues who noted transient elevation in mood in several healthy subjects who had received single pulses of TMS to the motor cortex.[6] Several small open studies followed that suggested that low frequency rTMS over the vertex may have antidepressant effects. Since the mid 1990s most interest has focussed on high-frequency rTMS applied to the left dorsolateral prefrontal cortex (LDLPFC), a region reported to be underactive in depression. To date there have been approximately 30 randomized trials of real and placebo rTMS in depression. In addition there have been several published meta-analyses including a Cochrane review.[7] This reviewed 16 trials, 14 of which were suitable for quantitative analysis. They found that high-frequency rTMS to the LDLPFC and low-frequency rTMS to the right dorsolateral prefrontal cortex (RDLPFC) were both superior to sham treatment but only for one measure (the Hamilton Depression Rating Scale) and at one time point (immediately after 2 weeks of treatment). The difference between real and sham was not large, leading to the conclusion that at this stage there was not strong evidence to support the use of rTMS as an antidepressant therapy.

There has been considerable heterogeneity between studies. Nearly all the trials have comprised patients with major depressive disorder defined using DSM-IV criteria. However there has been considerable variability with respect to pharmacotherapy received, with some trials specifying treatment resistance (variously defined) and some specifying medication-free participants. In two of the studies patients were started on antidepressant treatment either shortly before or simultaneously with the rTMS treatment.

The choice of appropriate sham condition is an important methodological consideration. There are two main approaches. Most studies have relied on tilting the active coil (usually through 45° or 90° with one or both wings of the coil touching the scalp). However intracerebral voltage measurements in a rhesus monkey have shown that, depending on how the coil is tilted, sham conditions obtained by coil tilting can induce voltages in the brain to levels only 24 per cent below active rTMS.[8] The fact that some 'sham' coils produce significant cortical stimulation may account for some of the benefit seen in those receiving sham stimulation and may underestimate the difference between real and placebo treatment. The other approach is to use a specially designed placebo coil. This looks identical to the real coil and makes the same noise but does not cause any cortical stimulation. However, neither does it cause sensation to the scalp, meaning that subject blinding may still be less effective. Indeed the problem of maintaining blinding in studies with rTMS continues to be a major methodological issue.

Most studies have given high-frequency rTMS to the LDLPFC, probably as a result of the positive early studies when this area was stimulated. Several investigators have used low-frequency rTMS to the RDLPFC. Low-frequency rTMS is much less likely to induce a seizure and is probably better tolerated by patients. Since slow rTMS has an inhibitory effect in contrast to the excitatory effect of fast rTMS and since there is considerable evidence that the left and right hemispheres have contrasting functions in regulating mood, it could be speculated that slow rTMS to the right cortex may have a similar effect to fast rTMS to the left.

Studies of high-frequency rTMS in depression have generally used stimulation frequencies of 5 to 20 Hz. There is a suggestion from animal studies that higher frequency stimulation may have a greater antidepressant effect but so far the numbers of subjects in human studies have been too low to show if a difference in effect of varying stimulation frequency exists. Likewise, the optimal stimulus intensity, length of treatment course, and total number of stimulations is not yet clear from the published data. However, longer trials with an increase number of stimulations appear to make little difference.[9]

Most of the rTMS studies in depression have been small, the largest until recently having 70 patients. However, recently a much larger industry-sponsored (Neuronetics) trial submitted their findings to the US Food and Drug Administration (FDA), seeking licensing approval for an rTMS device. In this study 301 patients were randomized to real or sham rTMS. Participants received 10 Hz rTMS of the left dorsolateral prefrontal cortex, 3000 pulses per day for 20 days. Although there was a marginal difference between the groups in favour of rTMS at the end of treatment, there was no significant group difference on an intention-to-treat analysis of the primary outcome measure (Montgomery-Åsberg Depression Rating Scale). In January 2007 the FDA Neurological Devices panel considered Neuronetic's application to have its rTMS equipment licensed for therapeutic use. The panel felt that there was insufficient evidence to support its efficacy. The final FDA decision is expected in summer 2007 (website: http://www.fda.gov/cdrh/panel/summary/neuro-012607.html, accessed: 5 June 2007).

In the United Kingdom the National Institute for Clinical Excellence (NICE) has issued recommendations stating 'Current evidence suggests there are no major safety concerns associated with transcranial magnetic stimulation for severe depression but there is no evidence that the procedure has clinically useful efficacy' (website: http://www.nice.org.uk/article.aspx?o=ip346consultation, accessed: 5 June 2007).

Comparisons with ECT

In addition to comparisons with placebo treatment, rTMS has also been directly compared with ECT in several studies. While ECT is the most effective treatment for severe depression in the short-term its use is limited by several issues, including acceptability to patients, the requirement to be anaesthetized, and the occurrence of side-effects, particularly cognitive side-effects. rTMS could be a potential alternative if it proved effective. In total there have been six published randomized controlled trials to date comparing ECT and rTMS. These trials have all had relatively small numbers of

patients, particularly when compared with trials of antidepressant medications. They have either shown rTMS to be less effective or not statistically different from ECT. The most recent and largest trial to date included 46 patients and compared 3 weeks of treatment with rTMS to a course of ECT.[10] The mean reduction in the Hamilton Depression Rating Scale achieved at the end of treatment was 14.1 points in the ECT group, compared with 5.4 points for the rTMS group, translating into a mean percentage reductions from baseline of 58 and 22 per cent, respectively. Overall, ECT was shown to be substantially more effective as a short-term treatment of depression than rTMS.

rTMS and schizophrenia

While most studies of rTMS within psychiatry have focussed on depression, there has been a growing interest in using rTMS as a possible treatment for schizophrenia. It has been used to treat both auditory hallucinations and to alleviate negative symptoms of schizophrenia.

Auditory hallucinations occur in approximately 70 per cent of patients with schizophrenia and in about a quarter of cases respond poorly if at all to antipsychotic medication. Recent advances in neuroimaging have enabled measurement of neural activity while hallucinations are being experienced and it has been demonstrated that auditory hallucinations are associated with activation in a number of brain areas, including the temporal cortex bilaterally. This area has been targeted in several rTMS studies using slow rTMS to reduce excitability.

The first account of using rTMS to treat auditory hallucinations reported improvement in the severity of hallucinations of three patients with schizophrenia who had 40 min/day of 1 Hz rTMS over 4 days.[11] There have now been 15 published treatment studies of rTMS targeting auditory hallucinations in schizophrenia. Ten sham-controlled trials (involving 212 patients) were included in a recent meta-analysis[12] which concluded that overall rTMS was significantly better than sham stimulation in the treatment of auditory hallucinations.

Negative symptoms of schizophrenia include alogia, avolition, anhedonia, and affective flattening and are associated with attentional impairment and executive dysfunction. Negative symptoms are often resistant to neuroleptic medication and are associated with poor clinical outcome. There is increasing evidence that negative symptoms are related to reduced cortical activation, particularly involving the left prefrontal cortex. Therefore one treatment approach has been to attempt to increase activation within this region. There have been four published randomized controlled studies comparing real and sham rTMS of the left dorsolateral prefrontal cortex to target negative symptoms of schizophrenia, of which three found no difference between real and sham treatments and one found 2 weeks of high-frequency rTMS significantly improved negative symptoms. Novak *et al.*[13] additionally performed a battery of neuropsychological tests and follow-up patients for 6 weeks after treatment but found no significant differences between treatment groups at either time point for primary or secondary outcome measures. The most recent study[14] did not provide evidence that rTMS to the DLPFC improved negative symptoms of schizophrenia in patients with prominent negative symptoms but did suggest that rTMS may improve cognitive functioning in this patient group, at least in the short-term. However larger studies with longer periods of follow-up will be required to further examine this preliminary finding.

rTMS and obsessive–compulsive disorder

There have been several studies that have attempted to treat symptoms of obsessive–compulsive disorder (OCD) by modulating activity in prefrontal and motor circuits using rTMS. The earliest blinded trial of rTMS for OCD included 12 patients and found that a single session of right prefrontal high-frequency cortical stimulation significantly decreased compulsive urges for over 8 h.[15] Obsessive thoughts did not change significantly. This study suggested that rTMS may be a useful probe of neuronal circuitry associated with symptoms of OCD. However a number of subsequent studies have failed to replicate these findings. A Cochrane review in 2003 examined three randomized controlled trials and concluded that there were currently insufficient data to draw conclusions about the efficacy of rTMS in the treatment of OCD.[16]

The most recent, and largest trial of TMS in OCD to date randomly allocated 33 patients with OCD to receive 10 sessions of either active or sham low-frequency (1 Hz) rTMS over the LDLPFC.[17] This study did not demonstrate any difference between real and sham treatments.

rTMS and other neuropsychiatric disorders

rTMS has been postulated as a potential treatment in a variety of other neuropsychiatric disorders. There is emerging evidence that it may improve some of the motor symptoms of Parkinson's disease. In a recent study six daily sessions of high-frequency rTMS were given to 55 unmedicated patients with Parkinson's disease.[18] Patients received either 10 Hz or 25 Hz rTMS bilaterally to the motor cortex arm and leg areas or to the occipital cortex (control group). It was found that stimulation to the motor areas improved all measures, e.g. walking time, key-tapping speed, and self-assessment and that 25 Hz stimulation yielded greater improvement than 10 Hz. The effect was sustained for a month after treatment and restored by further booster sessions. The authors concluded that 25 Hz rTMS can lead to cumulative and long-lasting benefits on motor performance in Parkinson's disease.

A recent study explored the effect of low-frequency rTMS to the left temperoparietal region on chronic tinnitus.[19] Patients received 1200 stimuli per day for 5 days of either real or placebo treatment in a randomized controlled crossover trial. Overall active rTMS induced a transient, but significant improvement in the symptoms of tinnitus.

rTMS has also been used in an attempt to provide relief from chronic neuropathic pain.[20] A recent review summarized that high-frequency rTMS to the motor cortex is able to produce pain relief but that the effect is brief and that research in this area is required.

rTMS has also been used as a probe of neuronal circuits in dementia and in attention deficit hyperactivity disorder although any possible therapeutic role for it in these conditions appears some way off.

Summary

rTMS has an increasing role as a useful investigational tool for probing neuronal circuitry in a variety of neuropsychiatric disorders. However its therapeutic value is at present less certain.

The antidepressant efficacy of rTMS has now been investigated for over 15 years and despite initial early enthusiasm there is still not clear evidence for its usefulness as a treatment in depression, reflected in the recent FDA and NICE decisions. Further research is required to identify specific brain regions in specific conditions that may be appropriate targets for treatment with rTMS, allowing tailoring of treatments for individual patients. The recent development of neuronavigational techniques using MRI imaging should aid treatment site localization. Other future research should be directed at establishing optimal rTMS parameters, e.g. the intensity, frequency and number of treatments.

Further information

http://www.ists.unibe.ch/ is the homepage of the International Society for Transcranial Stimulation. It also provides links to other websites of interest.

George, M.S. and Belmaker, R.H. (eds.) (2007). *Transcranial magnetic stimulation in clinical psychiatry*, p. 289. American Psychiatric Publishing. Arlington, VA. Provides a useful summary of the theoretical and practical aspects of TMS research.

Lisanby, S.H. (ed.) (2004). *Brain stimulation in psychiatric treatment*, p. 153. American Psychiatric Publishing, Washington, DC. Reviews transcranial magnetic stimulation and also other brain stimulation techniques including deep brain stimulation, magnetic seizure therapy and vagus nerve stimulation.

References

1. Barker, A.T., Jalinous, R., and Freeston, I.L. (1985). Non-invasive magnetic stimulation of human motor cortex. *Lancet*, **1**, 1106–7.
2. Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature*, **406**,147–50.
3. Speer, A.M., Kimbrell, T.A., Wassermann, E.M., *et al.* (2000). Opposite effects of high and low frequency rTMS on regional brain activity in depressed patients. *Biological Psychiatry*, **48**, 1133–41.
4. Strafella, A.P., Paus, T., Barrett, J., *et al.* (2001). Repetitive transcranial magnetic stimulation of the human prefrontal cortex induces dopamine release in the caudate nucleus. *The Journal of Neuroscience*, **21**, RC157.
5. Wassermann, E.M. (1998). Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the international workshop on the safety of repetitive transcranial magnetic stimulation, June 5–7, 1996. *Electroencephalography and Clinical Neurophysiology*, **108**, 1–16.
6. Bickford, R.G., Guidi, M., Fortesque, P., *et al.* (1987). Magnetic stimulation of human peripheral nerve and brain: response enhancement by combined magnetoelectrical technique. *Neurosurgery*, **20**, 110–6.
7. Martin, J.L., Barbanoj, M.J., Schlaepfer, T.E., *et al.* (2002). Transcranial magnetic stimulation for treating depression. *Cochrane Database of Systematic Reviews*, **2**, CD003493.
8. Lisanby, S.H., Gutman, D., Luber, B., *et al.* (2001). Sham TMS: intracerebral measurement of the induced electrical field and the induction of motor-evoked potentials. *Biological Psychiatry*, **49**, 460–3.
9. Loo, C.K., Mitchell, P.B., McFarquhar, T.F., *et al.* (2007). A sham-controlled trial of the efficacy and safety of twice-daily rTMS in major depression. *Psychological Medicine*, **37**, 341–9.
10. Eranti, S., Mogg, A., Pluck, G., *et al.* (2007). A randomized, controlled trial with 6-month follow-up of repetitive transcranial magnetic stimulation and electroconvulsive therapy for severe depression. *The American Journal of Psychiatry*, **164**, 73–81.
11. Hoffman, R.E., Boutros, N.N., Berman, R.M., *et al.* (1999). Transcranial magnetic stimulation of left temporoparietal cortex in three patients reporting hallucinated. "voices" *Biological Psychiatry*, **46**, 130–2.
12. Aleman, A., Sommer, I.E., and Kahn, R.S. (2007). Efficacy of slow repetitive transcranial magnetic stimulation in the treatment of resistant auditory hallucinations in schizophrenia: a meta-analysis. *The Journal of Clinical Psychiatry*, **68**, 416–21.
13. Novak, T., Horacek, J., Mohr, P., *et al.* (2006). The double-blind sham-controlled study of high-frequency rTMS (20Hz) for negative symptoms in schizophrenia: negative results. *Neuroendocrinology Letters*, **27**, 209–13.
14. Mogg, A., Purvis, R., Eranti, S., *et al.* (2007). Repetitive transcranial magnetic stimulation for negative symptoms of schizophrenia: a randomized controlled pilot study. *Schizophrenia Research*, **93**, 221–8.
15. Greenberg, B.D., George, M.S., Martin, J.D., *et al.* (1997). Effect of prefrontal repetitive transcranial magnetic stimulation in obsessive-compulsive disorder: a preliminary study. *The American Journal of Psychiatry*, **154**, 867–9.
16. Martin, J.L., Barbanoj, M.J., Perez, V., *et al.* (2003). Transcranial magnetic stimulation for the treatment of obsessive-compulsive disorder. *Cochrane Database of Systematic Reviews*, **3**, CD003387.
17. Prasko, J., Paskova, B., Zalesky, R., *et al.* (2006). The effect of repetitive transcranial magnetic stimulation (rTMS) on symptoms in obsessive-compulsive disorder. A randomized, double blind, sham controlled study. *Neuroendocrinology Letters*, **27**, 327–32.
18. Khedr, E.M., Rothwell, J.C., Shawky, O.A., *et al.* (2006). Effect of daily repetitive transcranial magnetic stimulation on motor performance in Parkinson's disease. *Movement Disorders*, **12**, 2201–5.
19. Rossi, S., De Capua, A., Ulivelli, M., *et al.* (2007). Effects of repetitive transcranial magnetic stimulation on chronic tinnitus. A randomised, cross over, double blind, placebo-controlled study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **78**, 857–63.
20. Leo, R.J and Latif, T. (2007). Repetitive transcranial magnetic stimulation (rTMS) in experimentally induced and chronic neuropathic pain: a review. *The Journal of Pain*, **8**, 453–9.

6.2.10.4 Neurosurgery for psychiatric disorders

Keith Matthews and David Christmas

Ablative neurosurgery

Definition

Historical definitions of Neurosurgery for Mental Disorder (NMD), previously known as 'psychosurgery', have either made distinctions between neurosurgery for psychiatric or 'psychological' illness and disorders assumed to have a clearer 'biological' origin (e.g. epilepsy, Parkinsons' disease); or, have emphasized control of behaviour as a therapeutic objective rather than the control of symptoms. A more recent definition, and the one used throughout this chapter is that provided by the UK Royal College of Psychiatrists:[1]

> *A surgical procedure for the destruction of brain tissue for the purposes of alleviating specific mental disorders carried out by a stereotactic or other method capable of making an accurate placement of the lesion.*

Historical overview

The first attempt at treating psychiatric illness by surgical methods is commonly attributed to Gottlieb Burckhardt, a Swiss psychiatrist, who in 1888 performed 'temporal topectomy' on six patients who were most probably suffering from schizophrenia. His intention

was to sever the connections between the frontal lobes and the rest of the brain. Results were mixed: one patient was reported as improved; two were 'quieter'; and two showed no change. However, one patient died; another developed epilepsy; and a further had motor weakness. His results were met with a mixture of ridicule and hostility and he never again wrote on the subject.

In 1935, James Fulton and Carlyle Jacobsen operated on the frontal lobes of two chimpanzees named Becky and Lucy after first studying their responses to frustration in behavioural experiments. They found their behaviour dramatically changed after the surgery. Becky's previous agitated responses to frustration became more passive whilst Lucy was much more agitated. At a London meeting in 1935, they presented their findings to an audience which included Egas Moniz, a Portuguese neurologist.

Moniz teamed up with Almeida Lima, a neurosurgeon, and in a 30-min operation in November 1935, they performed frontal leucotomy on their first patient. The procedure first involved injecting alcohol into the white matter tracts of the frontal lobes, but they later would change to using an instrument of their own design, the leucotome, to extirpate 'cores' of tissue. In 1936, they published their report on the outcomes of 20 patients who were probably suffering from depression, panic disorder, and schizophrenia. One-third were better, one-third were worse, and one-third were unchanged.

Shortly after their paper was published, Walter Freeman, a US neurologist, wrote an enthusiastic review and quickly secured the collaboration of a neurosurgeon, James Watts. They modified the procedure slightly and began practising what Freeman termed bilateral frontal lobotomy. Over the next decade, Freeman became frustrated with the cumbersome requirements of a neurosurgical theatre and team. Adapting a technique first described in the 1930s, Freeman infamously developed the transorbital lobotomy in 1946. Notorious for the initial use of an ice-pick, the procedure involved forcing a tool (an 'orbitoclast') under the upper eyelid and through the base of the skull into the frontal lobes. Also known as the 'ice-pick lobotomy', the relative ease with which the procedure could be performed resulted in the widespread adoption of the technique throughout the United States and Europe. However, it was the indiscriminate overuse of such 'freehand' procedures and the associated adverse effects that occurred in many patients that led to public and professional antipathy towards neurosurgery for psychiatric illness, which peaked in the late 1950s. The introduction of chlorpromazine in 1954 also meant that for the first time there was a non-surgical treatment for schizophrenia.

Despite a reduction in the use of neurosurgery for mental disorder in the late 1950s and early 1960s, the development of stereotactic techniques (which had been demonstrated in 1908 by Horsley and Clarke and adapted for human use in 1947 by Spiegel and Wycis) meant that greater accuracy and greater consistency in neurosurgery could deliver better outcomes in selected patients. Procedures became more selectively and reliably targeted and lesions became more discrete.

Procedures

All NMD procedures have targeted one or more of three main regions: (i) fronto-limbic connections within the orbital or cingulate cortices; (ii) subcortical limbic circuitry; and (iii) limbic cortex, including the amygdala and cingulate cortex.

Four 'modern' stereotactic procedures have been described with only two remaining in regular usage in the Western World. Anterior capsulotomy is still performed in Cardiff (UK), Spain, Belgium, and Scandinavia, whilst anterior cingulotomy is the procedure of choice in Dundee (UK), Poland, South Korea, and North America.

(a) Subcaudate tractotomy (SST)

Developed in the United Kingdom by Geoffrey Knight in 1965, lesions were originally created using radioactive Yttrium^{90} rods. SST targets the white matter tracts of the 'substantia innominata' connecting the orbital cortex to limbic regions, and probably involved lesioning the nucleus accumbens.

(b) Anterior capsulotomy (ACAPS)

Described by Jean Talairach in 1949 and further developed by Lars Leksell for the treatment of chronic pain, ACAPS places lesions in the anterior limb of the internal capsule—a large white matter bundle connecting the frontal cortex with the thalamus and limbic structures. Lesions are generated using focused gamma radiation (gammacapsulotomy) or thermal damage (thermocapsulotomy). (See figure 6.2.10.4.1 for typical ACAPS lesions.)

(c) Anterior cingulotomy (ACING)

The cingulate gyrus was first proposed as a target by John Fulton in the late 1940s. Hugh Cairns, an English neurosurgeon and a friend of Fulton's, performed 'cingulectomy' in 1948. The less destructive 'cingulotomy' was first performed by Eldon Foltz and Lowell White in 1962 for the treatment of pain. (See figure 6.2.10.4.2 for typical ACING lesions.)

(d) Limbic leucotomy

First developed by Desmond Kelly in 1973, it represented a combination of cingulotomy and subcaudate tractotomy lesions.

Indications

During the 1960s and 1970s, a variety of other operations were explored as treatments for hypersexuality, aggression, and criminality. These included hypothalamotomy and amygdalotomy. Reports of outcomes from such interventions were often favourable, but interpretation has been complicated by issues of diagnosis, patient selection, and assessment. Such clinical presentations are not considered appropriate indications for surgery today.

Three main indications exist for modern NMD: obsessive-compulsive disorder (OCD); anxiety disorders; and major depression. Only individuals who have experienced chronic, disabling symptoms that have failed to respond after the diligent pursuit of available treatments (pharmacological and psychological) should be considered for ablative neurosurgery.

There are few absolute contraindications to NMD. There is no evidence to support the use of NMD as a treatment for eating disorders, schizophrenia, or personality disorders. However, where these exist as significant comorbid conditions alongside depression or OCD, these do not represent absolute contraindications and careful consideration is required.

Ethical considerations

One of the most persistent concerns for the public and for health professionals is that NMD is used to treat patients in the absence of informed consent. Although the absence of informed consent was likely to be an issue for early procedures, to the best of our knowledge,

all contemporary centres performing NMD today insist upon the patient's ability to give informed consent. In Scotland, the Mental Welfare Commission must authorize any proposed NMD as being in the patient's best interests and must confirm that the patient is capable of providing informed consent. In England and Wales, the Mental Health Act Commission has a duty to set-up multi-disciplinary panels to authorize NMD for consenting patients under Section 57 of the Mental Health Act, 1983. Similarly, in most other countries where the procedure is available, the procedure can only go ahead with the approval of an independent review board.

Criteria for NMD

General criteria for suitability show little variation between centres. Key inclusion and exclusion criteria for NMD in Dundee are shown below in Table 6.2.10.4.1.

Outcomes from NMD

Whilst placebo-controlled trials are considered the ideal assessment for intervention trials, they have frequently been described as unethical in the case of NMD. Despite this, there have been three isolated double-blind trials, involving a total of 6 patients. Despite suggestions of non-response in all cases, follow-up was brief and there is inadequate detail to make informed judgements of outcome.

(a) Combined outcomes for different procedures

Comparisons across studies are of limited value due to differences in procedure, patient characteristics, and the use of different rating scales. However, Spangler *et al.*[2] reviewed outcomes from different procedures and for different indications, defining a positive outcome as being a score of 1 or 2 on the Clinical Global Impression (Improvement) scale.[(2)] Positive outcome rates were: ACAPS (67 per cent); ACING (61 per cent); SST (37 per cent); and limbic leucotomy (67 per cent). The most effective procedures for affective disorder and OCD (respectively) were limbic leucotomy followed by ACING. The least effective procedure for both disorders (and overall) was SST.

Table 6.2.10.4.1 Inclusion and exclusion criteria for NMD

Inclusion criteria	Exclusion criteria
1. Age ≥20 years	1. Age <20 years
2. *Legal status*: both formal and informal patients can be considered	2. Failure to fulfil ICD-10 criteria for a suitable indication
3. *ICD-10 diagnosis of*: severe depressive episode; recurrent depressive disorder, current episode moderate to severe; bipolar affective disorder, current episode severe depression	3. *Primary diagnosis of*: substance misuse; organic brain syndrome; adult personality disorder; pervasive developmental disorder
4. *Duration of episode of illness*: minimum of 3 years, with at least 2 years of unremitting symptoms despite active treatment. Only in exceptional circumstances would a duration<5 years be considered	4. Absence of evidence of an adequate therapeutic trial of psychological treatment
5. *Consent*: the patient must be capable of providing sustained, informed consent	5. Absence of evidence of extensive trials of adequate pharmacological treatment

(b) Anxiety disorders

The crude rate of improvement following NMD (all procedures) for anxiety disorders ($n = 290$) is 77 per cent.[(1)] More recent reports of ACAPS would support a claim to effectiveness but this may be at the expense of significant adverse effects (apathy and dysexecutive symptoms).

(c) Obsessive-compulsive disorder

The combined rate of 'Completely Improved' or 'Improved' outcomes following SST is 52 per cent. More recent studies involving ACAPS or ACING report improvements on the Yale-Brown Obsessive-Compulsive scale (Y-BOCS) in the region of 30 per cent. Despite this relatively low figure, approximately 85 per cent of patients ($n = 478$) will have a marked or lesser improvement following NMD for OCD.[(1)]

(d) Depression

There is only one report of outcomes from ACAPS for depression. Herner (1961) described outcomes for 19 patients with a 'depressive state'.[(3)] Outcomes included: 'permanent improvement' (74 per cent); unchanged (5 per cent); and worse (5 per cent). In all 75 per cent experienced permanent side effects.

Spangler *et al.* (1996) reported that 53 per cent of those with affective disorder ($n = 10$) responded to ACING, with a 60 per cent response rate in unipolar depression.[(2)] Dougherty *et al.* (2003) reported a mean reduction in Beck Depression Inventory (BDI) score of 33 per cent in 13 patients following ACING.[(4)] With regards to limbic leucotomy, there are few studies looking at outcomes solely in depression but Mitchell-Heggs (1976) reported that 7 of 9 patients with depression improved after surgery.[(5)]

(e) Bipolar disorder

There are only two reports of NMD for bipolar disorder, both following SST.[(6,7)] Each involved small numbers ($n = 9$) but described improvements in cycle frequency with a greater effect on manic episodes than depression. Improved drug responsiveness was also alleged.

Mechanism of action

The mechanism of action of NMD is unknown, but almost all neurosurgical procedures involve lesioning white matter tracts connecting the prefrontal cortex with the thalamus, cingulate gyrus, and areas of the limbic system such as the amygdala and hippocampus. Neuroimaging studies have suggested that this circuitry is dysfunctional in depression and interrupting parts of these circuits may rectify emotional and cognitive processing within these brain areas. In the case of OCD, there is compelling evidence that symptoms arise from functional circuits connecting the frontal cortex, thalamus, and basal ganglia and that lesioning parts of this circuit may serve to eradicate many of the symptoms.

Adverse effects

(a) General adverse effects

Transient adverse effects such as headache are relatively common and tend to resolve in the first week after surgery. Post-operative confusion can occur in 3–10 per cent of patients with higher rates following SST. Incontinence is relatively uncommon with modern procedures, but the reported rates are: 1.1 per cent after SST;

5.5 per cent after ACING; and 9.5 per cent with limbic leucotomy. Apathy has been reported to occur in up to 24 per cent of patients following limbic leucotomy, but not all studies report its occurrence and this is likely to be a relatively high estimate. The incidence of weight gain varies greatly: 6.2 per cent after SST; 65.5 per cent after ACAPS; and 5.5–21.4 per cent after ACING. Similarly, there is wide variation in the reported rates of seizures: 1.6–3.3 per cent after SST; 0–7.7 per cent after ACAPS; 1–9 per cent after ACING; and 14.2 per cent after limbic leucotomy. Finally, suicide rates range from 1 per cent after SST to 12 per cent after ACING but these rates may reflect differing severities of illness.

(b) Effects on personality

It is certainly the case that early procedures such as leucotomy had marked effects on the personality and behaviour of large numbers of patients. However, with the advent of stereotactic procedures and more focused lesions, the effects on personality appear to be mild, sometimes even absent. It is acknowledged, however, that it is difficult to make robust appraisals of personality without assessment tools designed for such a purpose and in the context of symptom reduction in chronic illnesses.

Most published studies have reported normalization of personality traits following modern procedures such as ACING and ACAPS. In addition, there is a trend towards reductions in neuroticism and increases in extraversion. There are some recent reports of adverse effects on executive function following ACAPS for anxiety disorders[8] but many such studies lack preoperative assessments of personality making conclusions difficult to draw.

(c) Effects on neuropsychological function

As with personality changes, the detrimental effects of earlier procedures upon neuropsychological functioning were probably significant. However, the majority of studies reporting neuropsychological outcomes from ACAPS, ACING, and limbic leucotomy from the early 1970s onwards report either no deterioration on general measures (such as IQ, attention, memory) post-operatively or, more frequently, improvement. It is likely that improvements in performance are mediated through symptom reduction.

Vagus nerve stimulation (VNS)

Overview

The vagus nerve is the longest of the 12 cranial nerves and 80 per cent of its fibres are sensory afferents. These fibres terminate in the nucleus tractus solitarius, sending ascending fibres to the forebrain via the locus coeruleus and parabrachial nucleus. The vagus nerve, therefore, provides an access route to modify information which is processed in brain regions involved in mood regulation.

VNS involves the subcutaneous implantation of a programmable pulse generator in a location similar to a cardiac pacemaker. Electrodes connect the generator to the left cervical portion of the vagus nerve. Stimulation is delivered in an intermittent pattern (typically 30 s every 5 min) but parameters are changed using a palmtop computer and a programming wand held over the pulse generator.

VNS was first used to treat epilepsy in 1988 and became available for the treatment of refractory partial seizures in 1994. The first trials in depression began in 1998.

Outcomes in depression

There are a number of short, open trials of VNS which typically report a 3-month response rate (≥50 per cent reduction in the 24-item Hamilton Rating scale for Depression; $HRSD_{24}$) of 30–40 per cent, and a remission rate ($HRSD_{24} \leq 9$) of approximately 15 per cent.

Larger, 12-month trials have demonstrated 12-month response rates of 27.2–46 per cent and remission rates of 15.8–29 per cent.[9,10] In a 12-month controlled comparison of VNS versus Treatment-As-Usual (TAU), George *et al.*[11] reported response rates of 27 per cent for VNS + TAU versus 13 per cent for TAU. Such improvements appear to be maintained at 2 years, with response rates of 42 per cent and remission rates of 22 per cent.[12] In the only randomized, controlled trial of VNS, Rush *et al.*[13] reported 10-week response rates on the $HRSD_{24}$ of 15.2 per cent in the VNS group versus 10.0 per cent in the placebo group, changes which were non-significant.[13] Despite positive results in uncontrolled trials, definitive evidence of efficacy remains elusive.

Adverse effects

Most adverse effects are related to stimulation, and in most people they are fairly mild and improve over time. Many can be managed by altering the stimulation parameters. In the initial stages, common effects are: hoarse voice (53 per cent); headache (23 per cent); neck pain (17 per cent); cough (13 per cent); and dyspnoea (17 per cent). At 12-months, the only adverse effect to persist at rates higher than 10 per cent is hoarse voice (21 per cent).[9]

Deep brain stimulation (DBS)

Overview

As with VNS, DBS has evolved from a treatment for neurological disorders to a putative intervention for psychiatric illness. The most

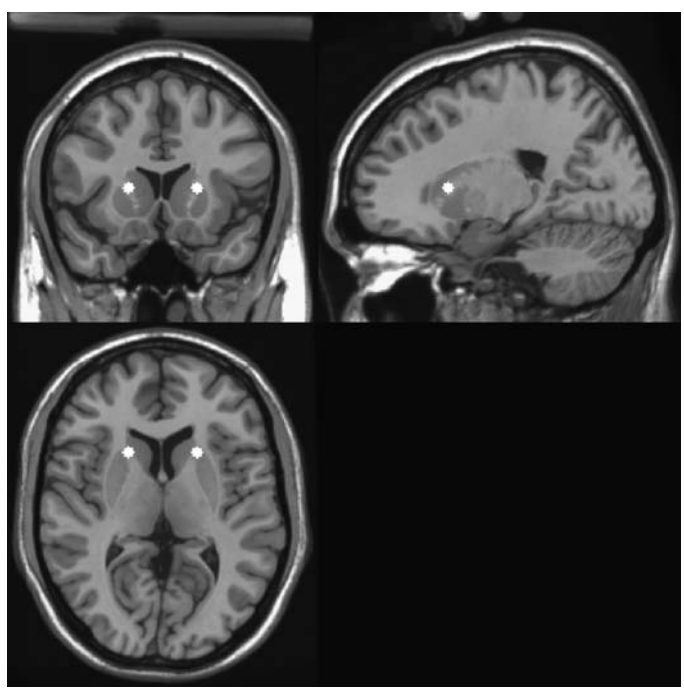


Fig. 6.2.10.4.1 Typical locations of anterior capsulotomy lesions superimposed upon normalized T1 MRI scan.
Lesions not to scale.

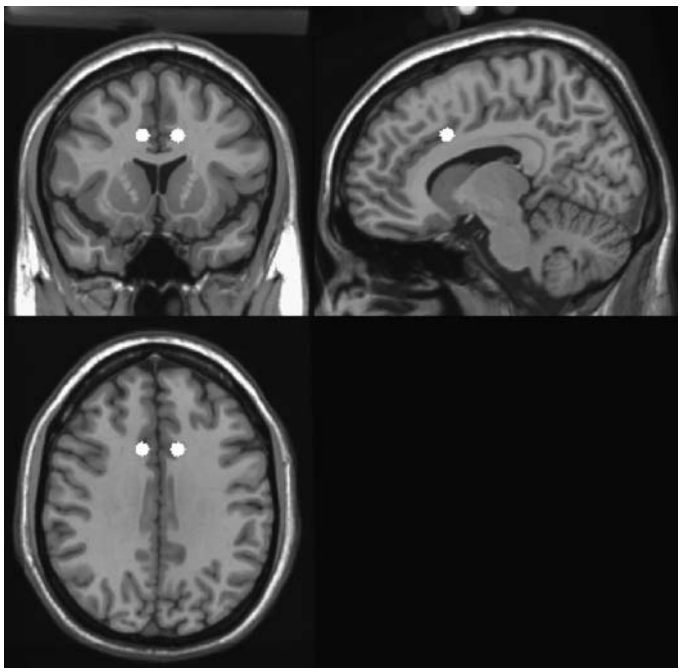


Fig. 6.2.10.4.2 Typical locations of anterior cingulotomy lesions superimposed upon normalized T1 MRI scan. Lesions not to scale.

effective targets for OCD and depression have yet to be determined but a number of possible locations for stimulation exist. However, DBS should be considered as an experimental treatment for both disorders.

The procedure involves the bilateral implantation of electrodes using stereotactic guidance with post-operative confirmation of location using MR scanning. The electrodes are connected to a generator typically implanted in the abdomen. Following surgery, the stimulation settings are programmed using immediate/short-term changes in symptoms as a guide.

DBS for obsessive-compulsive disorder

The most common target for DBS for OCD, thus far, has been ventral portion of the anterior limb of the internal capsule, the same site as ablative anterior capsulotomy. In a double-blind cross-over trial of anterior capsular stimulation in four patients Nuttin *et al*. (2003) reported reductions in symptoms of 36.8 per cent which were maintained after 21 months. In a small case series of four patients Abelson *et al.*[14] described marked improvements in one patient, with a lesser improvement in another. Greenberg *et al*. (2006) reported responses (≥35 per cent reduction in Y-BOCS score) in 4 out of 8 patients with DBS in the internal capsule.

Other proposed targets include the nucleus accumbens and the ventral caudate nucleus, but all targets may involve stimulation of a common anatomical area.

DBS for depression

In the only published report of DBS for depression, Mayberg *et al.*[15] stimulated the white matter tracts of the subgenual cingulate gyrus in six patients. Four patients were responders whilst two showed no change. Randomized on-off-on-off trials confirmed a stimulation-related improvement which was associated with a reduction in local cerebral blood flow in the subgenual cingulate and dorsolateral prefrontal cortex.

Adverse effects

Adverse effects that have been reported include: throbbing or buzzing sensations; nausea; and jaw tingling. A number of reports have described problems with battery life with the stimulators being replaced every 5 to 12 months. Battery failure has often been associated with a recurrence of symptoms over a few days which has been associated with marked depressive symptomatology and suicidal ideation. One study has reported a suicide, but commented that this was unrelated to stimulation. Numbers of reported cases are too small to determine if this is indeed the case.

Conclusions

Despite its chequered past, modern NMD bears little similarity to historical freehand procedures. Advances in neuroimaging mean that anatomical substrates for depression and OCD are being elucidated. Ablative procedures such as ACAPS and ACING are unlikely to undergo randomized controlled trials but prospective clinical audit suggests that such procedures may offer improvement to selected patients with treatment-refractory depression and OCD. Interventions such as VNS and DBS offer the possibility of double-blind testing but as yet there is insufficient evidence to suggest that such procedures offer greater effectiveness.

Further information

Freeman, C., Crossley, D., and Eccleston, D. (2000). *Neurosurgery for mental disorder. Report from the Neurosurgery Working Group of the Royal College of Psychiatrists.* Royal College of Psychiatrists, London.

Binder, D.K. and Iskandar, B.J. (2000). Modern neurosurgery for psychiatric disorders. *Neurosurgery*, **47**, 9–21.

Feldman, R.P., Alterman, R.L., and Goodrich, J.T. (2001). Contemporary psychosurgery and a look to the future. *Journal of Neurosurgery*, **95**, 944–56.

George, M.S., Rush, A.J., Sackeim, H.A., *et al.* (2003). Vagus nerve stimulation (VNS): utility in neuropsychiatric disorders. *International Journal of Neuropsychopharmacology*, **6**, 73–83.

Malone, D.A., Greenberg, B.D., and Rezai, A.R. (2004). The use of deep brain stimulation in psychiatric disorders. *Clinical Neuroscience Research*, **4**, 107–12.

References

1. Freeman, C., Crossley, D., and Eccleston, D. (2000). *Neurosurgery for mental disorder. Report from the Neurosurgery Working Group of the Royal College of Psychiatrists.* Royal College of Psychiatrists, London.
2. Spangler, W.J., Cosgrove, G.R., Ballantine, H.T. Jr., *et al.* (1996). Magnetic resonance image-guided stereotactic cingulotomy for intractable psychiatric disease. *Neurosurgery*, **38**, 1071–6.
3. Herner, T. (1961). Treatment of mental disorders with frontal stereotaxic thermo-lesions: a follow-up study of 116 cases. *Acta Psychiatrica Scandinavica*, **36**(Suppl. 158), 1–140.
4. Dougherty, D.D., Weiss, A.P., Cosgrove, G.R., *et al.* (2003). Cerebral metabolic correlates as potential predictors of response to anterior cingulotomy for treatment of major depression. *Journal of Neurosurgery*, **99**, 1010–7.
5. Mitchell-Heggs, N., Kelly, D., and Richardson, A. (1976). Stereotactic limbic leucotomy—a follow-up at 16 Months. *The British Journal of Psychiatry*, **128**, 226–40.

6. Lovett, L.M. and Shaw, D.M. (1987). Outcome in bipolar affective disorder after stereotactic tractotomy. *The British Journal of Psychiatry*, **151**, 113–16.
7. Poynton, A., Bridges, P.K., and Bartlett, J.R. (1988). Resistant bipolar affective disorder treated by stereotactic subcaudate tractotomy. *The British Journal of Psychiatry*, **152**, 354–8.
8. Rück, C., Andreewitch, S., Flyckt, K., *et al.* (2003). Capsulotomy for refractory anxiety disorders: long-term follow-up of 26 patients. *The American Journal of Psychiatry*, **160**, 513–21.
9. Marangell, L.B., Rush, A.J., George, M.S., *et al.* (2002). Vagus nerve stimulation (VNS) for major depressive episodes: one year outcomes. *Biological Psychiatry*, **51**, 280–7.
10. Rush, A.J., Sackeim, H.A., Marangell, L.B., *et al.* (2005). Effects of 12 months of vagus nerve stimulation in treatment-resistant depression: a naturalistic study. *Biological Psychiatry*, **58**, 355–63.
11. George, M.S., Rush, A.J., Marangell, L.B., *et al.* (2005). A one-year comparison of vagus nerve stimulation with treatment as usual for treatment-resistant depression. *Biological Psychiatry*, **58**, 364–73.
12. Nahas, Z., Marangell, L.B., Husain, M.M., *et al.* (2005). Two-year outcome of vagus nerve stimulation (VNS) for treatment of major depressive episodes. *The Journal of Clinical Psychiatry*, **66**, 1097–104.
13. Rush, A.J., Marangell, L.B., Sackeim, H.A., *et al.* (2005). Vagus nerve stimulation for treatment-resistant depression: a randomized, controlled acute phase trial. *Biological Psychiatry*, **58**, 347–54.
14. Abelson, J.L., Curtis, G.C., Sagher, O., *et al.* (2005). Deep brain stimulation for refractory obsessive-compulsive disorder. *Biological Psychiatry*, **57**, 510–6.
15. Mayberg, H.S., Lozano, A.M., Voon, V., *et al.* (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, **45**, 651–60.

6.3

Psychological treatments

Contents

6.3.1 Counselling

Diana Sanders

Introduction

People seek counselling for many reasons. Sometimes those who have had no previous need for mental health services are literally stopped in their tracks by life events—illness, family breakdown, intolerable stresses. People with long-term difficulties may turn to counselling when they feel the statutory services are not able to meet their needs, or as an adjunct to health care provision. With greater social mobility and the separation of family members, counselling increasingly provides the care and support previously offered within local communities. The provision and acceptability of counselling is on the increase. Counselling is possibly the most commonly delivered form of psychological therapy[1,2] and the British Association for Counsellors and Psychotherapists have over 30 000 members, with equivalent numbers in other countries. Professional training programmes in counselling have mushroomed in response to demand. Counsellors are found in many statutory and voluntary settings—mental health, primary care and medical settings, workplaces, drug and alcohol services, voluntary and charitable organizations, trauma services, and educational settings—as well as in private practice.

But what exactly is counselling? What do counsellors do? Is counselling the same as psychotherapy? And, is it an effective form of treatment? Although counselling is a major growth area within mental health, it can be difficult for consumers and purchasers of counselling services to know what kind of counselling and counsellor to use, with lack of clarity about what works for whom. There are many different models of counselling, types of counsellor and many different training courses. It is difficult to make clear distinctions between counselling and psychotherapy. Much of the work of counsellors has not historically been amenable to standard methods of evaluation, and research is relatively new. Currently there is no statutory regulation for the term 'counsellor', which means that people are able to practise as counsellors without registration or accreditation. By definition, people who seek counselling are likely to be vulnerable, and the issue of public protection is paramount.

The aim of this chapter is to clarify these issues and examine the place of counselling in psychiatry. The chapter begins by looking at the definition of counselling, and how counselling is both similar to, and distinct from, psychotherapy. The chapter goes on to look at the key features of counselling, and different models of counselling. Although counselling can and is used for many psychological difficulties, the chapter selects specific problems where there is evidence that it is an effective intervention: mild to moderate depression, adjustment difficulties, bereavement, trauma, and relationship problems. I then consider counselling in different settings, again selecting a few which illustrate the work of counsellors—primary care, mental health settings, student counselling, and the workplace—looking at the way counselling can be adapted according to the needs of the service. The chapter concludes by looking at issues of training, quality, and standards, commenting on the need for the control of an ever-developing profession without loss of the growing availability of effective counselling services to those in need.

Defining counselling

No single definition of either counselling or psychotherapy exists in spite of many attempts in Britain, North America, and elsewhere to arrive at one.[3] Currently, neither the British Association of Counselling and Psychotherapy (BACP) nor the American Counseling Association has either proprietary rights of the terms or even official definitions, although, as discussed below, the move towards statutory regulation for counselling and psychotherapy may ensure greater clarity for practitioners and consumers.

At its broadest, counselling is conceptualized as a way of helping or assisting others to make their own adjustment and decisions in the face of life problems. Counselling aims to offer a safe relationship within which the individual can explore personal difficulties and, through developing a deeper understanding of themselves, move towards change. The Department of Health defines counselling as . . . a *form of psychological therapy that gives individuals an opportunity to explore, discover, and clarify ways of living more resourcefully, with a greater sense of well being. Counselling may be concerned with addressing and resolving specific problems, making decisions, coping with crises, working through conflict, or improving relationships with others.*[4]

In contrast with other forms of psychological therapy, where the focus may be on treating specific problems, many counselling models give equal if not more weight to the *process* of change. The journey through which a client goes—to greater understanding, awareness, and resolution—is as important as the outcome. Counsellors practise within all therapeutic approaches, strongly influenced by humanistic, experiential, and psychodynamic principles. They tend not to link their work to diagnostic categories, preferring to see each client as an individual, and using an approach matching the client's needs rather than diagnosis. Counsellors may define themselves by the model they practise—for example, humanistic or psychodynamic—and/or by the type of problems they work with, such as bereavement or relationship counsellors.

Counselling may also be defined in terms of *key elements* and *goals*,[5] as shown in Table 6.3.1.1. Again, the key elements and goals illustrate the variation within counselling. For example, what is meant by counselling ranges from providing a safe arena for people to gain understanding and insight, to offering more direction and guidance leading to decision-making and problem-solving.

Table 6.3.1.1 The key elements and goals of counselling (Reproduced from Feltham, C. What are counselling and psychotherapy? In The Sage Handbook of Counselling and Psychotherapy, (eds. C. Feltham and I. Horton), pp. 3–10, copyright 2006, Sage Publications.)

The key elements of counselling	The goals of counselling
◆ Listening and talking methods of addressing psychological and psychosomatic problems and change ◆ An unstructured and non-directive form of therapy, using the therapeutic relationship as an active ingredient in promoting change ◆ Counselling operates largely without medication or other physical interventions ◆ Counselling may be concerned not only with mental health but with social, spiritual, philosophical and other aspects of living ◆ Professional forms of counselling are based on formal training, accreditation and on-going supervision and professional development	◆ Support, psycho-education and guidance ◆ Insight and understanding ◆ Self actualization and personality change ◆ Adjustment, symptom reduction and 'cure' ◆ Problem-solving and decision-making ◆ Crisis intervention and management ◆ Risk management (e.g. genetic counselling)

Counselling and psychotherapy

Much of the above can also be applied to psychotherapy and parts of this chapter do indeed overlap significantly with psychotherapy. There may well be variation between countries in what is defined as counselling or psychotherapy. Counselling and psychotherapy each have distinct features including different historical roots. *Psychotherapy* arose from the seminal works of Freud in the late nineteenth and twentieth centuries, and in the past, psychotherapists tended to offer a long-term psychodynamic approach. Now, however, psychotherapy also includes interpersonal, humanistic, and cognitive models. In the United States, *counselling* was originally linked with vocational guidance, personnel management, and the workplace,[5] and as such was much more advisory and directive than the analytic processes of psychotherapy. Carl Rogers, the founder of non-directive counselling in the 1940s, initiated the movement away from practical guidance and problem-solving towards collaborative and person-centred models, forming the basis of counselling today.

Differences between psychotherapy and counselling tend to relate more to the individual psychotherapist's or counsellor's training and interests and to the setting in which they work, rather than to any intrinsic difference in the two activities. In medical and mental health settings, psychotherapists are more likely to work with patients with severe psychological disorders, offering long-term therapy, whereas counsellors may concentrate on difficulties amenable to short-term work—mild to moderate psychological disorders, relationship difficulties, or bereavement. Counsellors who work for voluntary agencies or in educational settings such as schools and colleges usually concentrate more on the 'everyday' problems and difficulties of life than on severe psychological disorders, although agencies such as MIND, Alcoholics Anonymous, or

Narcotics Anonymous offer counselling to people with serious mental illnesses. In private practice, however, a counsellor's work will overlap with that of a psychotherapist.

In a pragmatic vein, Feltham states that *practitioners and the public stand to gain much more from the assumption of commonality than from spurious or infinitesimal distinctions . . . little is to be gained practically from further controversy about professional titles and distinctions*.[5] In 2000, the British Association for Counselling lent weight to greater rapprochement by becoming the British Association for Counselling and Psychotherapy. On a practical level, it is interesting that many recipients of 'talking therapies' other than counselling, such as psychotherapies, cognitive-behaviour therapy and problem-focused discussions with GPs, psychiatrists, or nurses, say that they have received *counselling*, reinforcing Feltham's plea that the issue of definition and distinction is academic rather than of practical value.

Counselling skills and counselling practice

Counselling skills are integral to the practice of psychiatry and all the 'helping professions', as basic ingredients of effective interviewing, accurate history-taking, diagnosis, and treatment-planning.[6] The skills of listening, summarizing, reflecting, checking, understanding, gaining rapport, and communicating enable other people to feel understood. They are essential for engagement and eliciting information, especially when the person is afraid, in pain, or mistrustful. The health worker's counselling skills may influence the patient's collaboration with an active participation in treatment, and thereby the outcome of a wide range of medical and even surgical treatments. Many helping and health professionals such as social workers, occupational therapists, probation officers, and speech and communication therapists use counselling skills as an integral part of their work but would not be seen as primarily counsellors.

In contrast, counsellors as *professionals*, who use counselling as a specific intervention, work in many areas of mental health practice alongside psychotherapists, clinical and counselling psychologists, psychiatrists, psychiatric nurses, and social workers.[7] For professional counsellors, counselling skills are central to their work.

Counselling as a specific planned intervention in psychiatry can be differentiated into two broad and overlapping categories, defined by aims into *decision-making* and *treatment*:

- *Decision-making* is an important ingredient in many forms of therapeutic counselling but, conversely, some forms of decision-oriented counselling (e.g. genetic counselling) embody no explicit therapeutic intention.
- Counselling as a primary *treatment* for problems is used in the management of a range of conditions as an adjunct to other interventions including medication, as an integral component of a multimodal treatment method (e.g. crisis intervention), or as a specific treatment in its own right (e.g. for postnatal depression).

Counselling psychology

As well as professional counsellors, counselling psychologists have a particular role to play within counselling provision in mental health. The area of counselling psychology, now developing in the United Kingdom in line with other parts of the world, is a distinctive profession within applied psychology, which aims to foster the psychological development of the individual and help people develop more effective and fulfilled lives. It is based on the fundamental tenets of counselling, but in addition aims to integrate the application of psychological theory and research into its practice.[8] Counselling psychologists use a variety of therapeutic models, including person-centred, psychodynamic, and cognitive. Although the training of counsellors can be varied, as discussed below, counselling psychologists undergo standardized post-graduate doctorate training leading to chartered status within the British Psychological Society, or equivalent in other countries.

Is counselling an effective method of treatment?

Despite the proliferation of counsellors in many areas of medicine and psychiatry, counselling has tended to lag behind medicine and other health care professions in engaging in and promoting research to establish its effectiveness and efficacy.[9] The nature of counselling can mean that standard methods such as RCTs are not appropriate means of evaluation, whereas qualitative research methodology is better able to assess meaningful changes.[10] However, counselling as a profession is now engaging in better quality research, concentrating on outcomes in routine practice as well as qualitative analysis. New practice-based methods of evaluation, such as CORE (Clinical Outcomes in Routine Evaluation), and the aggregation of data across UK NHS counselling services can lead to national benchmarks. Methods of case-study research, and the development of measures of the client's perspective on psychological distress, PSYCHLOPS,[11] enable more client-focused research. Such emphasis on evidence-based practice will lead to more careful targeting of specific counselling approaches to specific problems,[3] clearer information for the public and will improve counselling's parity with other health care professions.

Currently, counselling has an image that it is more appropriate for people with mild to moderate difficulties. The Department of Health[4] recommends that counselling should not be the main intervention for people with severe and complex mental health problems or personality disorders. Patients who are adjusting to life events, illnesses, disabilities, or losses may benefit from brief therapies such as counselling. However, counsellors such as those in primary care and the voluntary sector are already offering an important and valued service. Although people with more serious and enduring mental health difficulties require primarily psychiatric and pharmacological intervention, offering emotional support, advice, and problem-solving can form an important, although under-researched, part of their care.

The core conditions of counselling and the therapeutic relationship

Counselling depends primarily on the interaction between the counsellor and client, what goes on in that interaction and the qualities of both client and counsellor. Carl Rogers'[12] definition of the conditions necessary for therapeutic change was a radical departure from traditional psychotherapeutic practice, in emphasizing the *qualities* and *attitudes* of the counsellor rather than specifying what the counsellor must *do*. His work led to the following as *necessary* and *sufficient* conditions for therapeutic change:

- The client is in a state of *incongruence*, being vulnerable or anxious.
- The therapist is *congruent* and *genuine* in the relationship with the client.

- The therapist experiences *unconditional positive regard* for the client.
- The therapist experiences an *empathic understanding* of the client's frame of reference or way of seeing things.
- The therapist feels *non-possessive warmth* towards the client.
- The client *perceives* the therapist's unconditional positive regard and empathic understanding.

These core conditions have been used and developed in many models of counselling and therapy; even therapies traditionally seen as more technical have always maintained their importance.[13] *Empathy*, for example, a core condition which is central to all good therapeutic relationships, enables clients to know that they are heard and understood. At its simplest, empathy is a simple restatement of someone else's words. At its richest, it involves ... *a fearless exploration of another's inner world, a sensing of meanings unspoken, a compassionate naming of pain ... the fullest empathy does not censor or discriminate. It sees the world as the other person sees it.*[14]

While Rogers took the view that such core conditions are both necessary and sufficient for therapeutic change to occur, other models of counselling have defined such conditions as necessary but not in themselves sufficient for change. However, the core conditions remain the bedrock upon which counselling is practised.

The therapeutic relationship

Across very diverse treatments, including cognitive and psychopharmacological,[15] measures of the strength of the relationship, or alliance, have been the strongest and most consistent process correlates of treatment outcome.[3,16] Clients who have strong alliances with their therapists tend to have better outcomes.

Although recognized as an essential component of change, different models have different conceptual and practical approaches to the relationship. Three examples illustrate the differences:

- Person-centred models take a here-and-now perspective, looking at the immediate interaction between client and counsellor. The client's perception of the therapist's empathy, unconditional positive regard, and congruence enables therapeutic change.
- Cognitive models regard the relationship as necessary but not sufficient for therapeutic change. The relationship is primarily collaborative, with an active, working bond formed between client and counsellor to facilitate the tasks of therapy.
- Psychodynamic models distinguish the real relationship between client and counsellor, and the transference relationship, consisting of both client transference and therapist counter transference. The working alliance therefore is only partly based in reality, also containing aspects of both parties' histories.

Counselling methods and techniques

There have always been many approaches to counselling and psychotherapy, and this diversity grew into a veritable 'multiverse' during which some authors estimated that there were over 400 brand therapies in existence. There are also different settings and agencies which offer counselling—clinics, institutes, health centres, or voluntary bodies, each with its own particular features. Within each model and setting, there are different formats of counselling including self-help materials on CD-rom and the Internet, as well as individual, couple, group, family, and organizational. Such a range can be confusing to potential clients and organizations, and the question of what works, for whom, and in which setting, has to be central in matching client, problem, therapy, and therapist.

Specific models of counselling are usually differentiated by a number of factors:

- Basic assumptions or philosophy
- Formal theory of human personality and development
- Clinical theory defining the goals, principles, and processes of change
- Therapeutic skills and techniques[17]

One useful distinction exists between *schools* of counselling and *theoretical approaches*.[18] A theoretical approach presents a single position regarding the theory and practice of counselling, whereas a school is a grouping of different theoretical approaches with common characteristics(see Table 6.3.1.2).

The three main schools are humanist-existential, psychodynamic, and cognitive behavioural. Humanistic-existential models will be described in detail, with briefer mention of psychodynamic and cognitive behavioural models, which are covered in other chapters. The section on methods also looks at the trend towards integration and eclecticism within counselling, whereby counsellors use a variety of methods and approaches adopted from different models. Although not clearly fitting into any one school, information-giving and *problem-solving* are counselling methods widely used in psychiatry, and are therefore described first.

Information-giving and problem-solving

Giving information is an important part of all medical and psychiatric practice, reflecting an open and collaborative approach to treatment, providing patients and their carers with the material necessary for informed decision-making. For example, for people with schizophrenia or those who misuse alcohol, the provision of information about the diagnosis, causes, and potential consequences of their condition is essential for mobilizing motivation and compliance with treatment. Giving information about the actions and potential side-effects of a prescribed medication enables people to

Table 6.3.1.2 Overview of counselling schools and main approaches (Reproduced from Nelson-Jones, R. *Theory and practice of counselling and therapy* (4th edn.), copyright 2006, Sage Publications.)

Psychodynamic school
Classical psychoanalysis (Sigmund Freud)
Analytical therapy (Carl Jung)
Humanistic-existential school
Person-centred therapy (Carl Rogers)
Gestalt therapy (Fritz Perls)
Transactional analysis (Eric Berne)
Existential therapy (Irvin Yalom and Rollo May)
Cognitive behavioural school
Behaviour therapy (Ivan Pavlov, BF Skinner and Joseph Wolpe)
Rational emotive behaviour therapy (Albert Ellis)
Cognitive therapy (Aaron Beck)
Multimodal therapy (Arnold Lazarus)

play an active role in pharmacological intervention. Information-giving is always crucial when communicating a diagnosis and fundamental to counselling for risk, as in genetic counselling, and to any intervention in which the individual is helped to make decisions.

Psycho-educative methods have a place in most models of counselling and psychotherapy, but have specific importance in problem-solving and cognitive behavioural models. For example, a psychologist or counsellor may describe to the client a psychological model of a specific condition, such as the cognitive model of panic, to help the client understand their particular symptoms.

Information-giving and psycho-education involves more than just giving information to a passive recipient. Wherever possible the individual's curiosity about their condition is promoted, encouraging them to ask questions and, when appropriate, to find their own answers. The Socratic method and guided learning are central to cognitive approaches. Information is not provided in a didactic fashion, but in response to the client's questions, as client and therapist are engaged in collaborative enquiry. Whatever the information given, the practitioner checks whether the client has understood the information and its meaning. Information-giving is rarely the endpoint of an intervention, serving instead as the basis for decision-making or continuing therapeutic work.

Problem-solving has been used and empirically validated as a specific treatment, particularly for depression, and is used by many cognitive behavioural and humanistic counsellors. Problem-solving forms a major part of brief solution-focused therapy.[19] From a problem-solving perspective, depression results from the interaction between negative life events, current problems, and deficient problem-solving abilities, and therefore facilitating solving problems is a means to alleviate depression.[20] Therapist and client work collaboratively to identify and prioritize key problem areas, break them down into specific manageable tasks, solve problems, and develop appropriate coping behaviours. The approach involves several stages:

- Identification and formulation of the client's problem(s)
- Setting clear and achievable goals
- Generation of alternatives for coping
- Selection and operationalization of a preferred solution
- Evaluation of progress, with further problem-solving as necessary

Research in the United Kingdom has shown that problem-solving delivered by general practitioners is as effective as pharmacological treatment for moderate and major depression in primary care.[21,22] The intervention can be extremely useful for clients who do not want or cannot tolerate pharmacological treatment and is recommended in NICE guidelines as a treatment for mild depression. It can be offered by counsellors, general practitioners, and nurses, and may be a means to improve treatment adherence for people with psychotic disorders, as part of a psycho-educational intervention including motivational interviewing.[23]

Brief solution-focused therapy developed from its roots in family therapy to applications in counselling, mental health, group work, education, drug and alcohol work, social work, and business. It is the preferred mode of working for counsellors in the workplace, given its brief and focused approach. The model arose from family therapists' observations that clients made significant changes when focusing on their preferred futures rather than on current problems. By articulating solutions, and building on existing skills and strengths, clients saw their problems in a different light and could effect change.

The 'miracle question' is a classic method of solution-focused therapy which is integrated into other models. The client is asked to think about and describe waking up one day to find that all problems have vanished. The counsellor explores the impact of the miracle on people and situations. The question enables the client to get into a problem-solving cognitive set, enabling identification of what needs to happen for the problems to change. The method has been studied in a range of client groups and settings, including with repeat offenders in the forensic service, and can produce positive outcomes.[19]

Humanistic and existential models

Humanistic and existential approaches include person-centred therapy, gestalt therapy, transactional analysis, and existential approaches. Of these, the person-centred model is the most well known, and the one that comes most readily to mind when describing the philosophy of counselling.

Client-centred, *or* ***person-centred*** as it is more often called, counselling originates from the work of Carl Rogers, whose emphasis on the recognition and empowerment of the help-seeker challenged the perceived authoritarianism of both the medical model and psychoanalysis. The model highlights respect for the person, and adopts the optimistic assumption that each person has an inner potential for healthy development and achievement, or 'self-actualization'. Person-centred approaches often use the analogy of a plant to describe the concept of growth and change. No one can make a plant grow, but if the plant is provided with the right conditions—water, light, soil, nutrients—then it will become the best plant it can be. Person-centred therapy assumes that people have an inbuilt motivation to change, and also have the skills necessary with which to effect changes. Rogers' model of counselling is non-directive. The counsellor's task is to create the core relationship conditions of empathy, warmth, unconditional positive regard, and genuineness, described above, in which the client's inner resources and potential will be unlocked, leading to the spontaneous resolution of problems and developmental growth.

The central features of person-centred counselling form the bedrock of other models of counselling, including cognitive approaches. Carl Rogers, in initiating the person-centred approach, has also had a wide influence in the helping professions—the term 'person-centred' is used frequently in policy documents and guidelines within health care organizations, as one of the standards of service and as a philosophy of health care.

While a non-directive and reflective approach has value, and may be useful for initial data-gathering and supportive work, caution must be applied to the use of Rogerian counselling in psychiatry. Resource constraints require practitioners to impose time limits on counselling, which therefore must be more focused and 'active'. Furthermore, very disturbed people may be unable to access an inner potential for spontaneous change and growth, implicit within the client-centred model. There are some for whom a reflective non-directive approach may be harmful, risking an overwhelming upsurge of avoided or forgotten memories of traumatic experiences without providing methods for coping with them.

Victims of childhood sexual abuse or other destructive experiences may be re-traumatized by unstructured reflective counselling.

It is likely that the person-centred approach will continue to form the basis of good counselling and psychotherapeutic practice regardless of the model used, with increased emphasis on more 'skills-based' approaches such as cognitive behavioural and other models that lend themselves more easily to measurement, structured working, and evidence-based practice.

Gestalt therapy was originated by Fritz Perls, who described his approach as dealing with the total existence of a person, rather than being primarily occupied with symptoms or character structure.[24] Gestalt therapy argues that the past is past and the future unknowable, therefore the focus of counselling should be the present moment—an approach, interestingly, espoused by the development of mindfulness in psychiatry and psychotherapy.[25] The goal of therapy is to put clients in touch with what they are thinking, feeling, and sensing, in the here and now, and how they restrict or limit themselves by continual focus on the past or future. Gestalt therapists regard the therapeutic relationship as a 'working' relationship, with client and counsellor taking responsibility for themselves. Attaining awareness is an essential aim within the relationship.

Gestalt therapy uses many techniques, including dream-work and psychodrama. The classic 'two chair' method of gestalt therapy enables clients to work with 'unfinished business' which may be influencing current problems. For example, a client with memories of a difficult relationship with a parent is encouraged to have a dialogue with the parent in the empty chair, to see both client and parent's point of view. The client may put themselves, metaphorically, in the empty chair, to enable greater understanding and acceptance of the self.

Research has only recently played a role in the development of gestalt theory and practice. Most of the studies concern the effectiveness of the two-chair method, an approach which is being integrated into other models, for example Greenberg's[26] emotion-focused psychotherapy, and cognitive approaches.[13]

Transactional analysis (TA) was founded by Canadian psychiatrist Eric Berne, and provides a theory of personality, child development, and psychopathology as well as a theory of counselling. The method assumes, as for other person-centred approaches, that people are born with a drive for growth and health—the 'I'm OK, you're OK' life position. TA characterizes the personality into three groups of 'ego states'—parent, adult, and child, each with behavioural, social, historical, and phenomenological aspects. Psychopathology arises from the repetition of unhelpful life scripts, or patterns of being, often learned early in life. Counselling enables the individual to identify and modify problematic patterns.

Very little research has been conducted into the effectiveness of TA as a therapy although many theoretical concepts and practical techniques have been assimilated into psychotherapy and counselling.[27] The method has also led to the concept of the 'reflective practitioner', a theme embodied by the BPS Division of Counselling Psychology.

Existential approaches originated in applied philosophy, and focus on helping people to come to terms with life in all its confusing complexity. Rather than curing people of pathology, the aim is to help people deal with the contradictions, dilemmas, and paradoxes of everyday existence.[28] Anxiety and depression, rather than to be avoided, are to be embraced and understood in order to live life to the full. The main method is conversational, enabling clients to confront rather than avoid the reality of situations.

Although existential approaches may sound idealized and unrealistic, much of existential therapy aims to help people build confidence and competence in tackling everyday problems. The methods and approaches have very little outcome research, because of the opposition of existential therapists to what is seen as the reductionist tendencies of research—i.e. what is effective in therapy is not open to evaluation using standard methodology. It may be that the approach offers a number of factors which can be usefully integrated into other, more evidence-based models, such as the focus on validating experience, creation of meaning to enable traumatic events to be processed, and authenticity.

Cognitive behavioural approaches

Cognitive behaviour therapy (CBT) is currently receiving excellent press internationally, and occupies a central place in the move towards evidence-based practice. NICE recommends CBT more often than other therapeutic approaches for many psychological problems. Despite its popularity and evidence-base, cognitive approaches have not been readily embraced by the counselling world.[13] The structured and focused approach, and use of techniques to promote change, rested uncomfortably with counsellors trained in client-centred approaches, and cognitive therapy was felt to pay insufficient attention to the therapeutic relationship and to the influence of past events on current problems. However, the last few years have seen a major change in the way cognitive therapy is being adopted within counselling, and a large proportion of counsellors integrate at least some of the approaches into their work.

The attraction of cognitive therapy to counsellors is increasing, with more overt focus placed on the therapeutic relationship, long-term approaches, and schema-focused work inherent in newer models, which enables counsellors to abandon their prejudices against CBT.[29] There is enormous scope for counsellors to adopt cognitive therapy in a more systematic and rigorous manner, particularly in light of the empirical evidence supporting its effectiveness and increasing demand for briefer interventions.[13] However, counsellors trained in different schools of counselling can be tempted to borrow specific methods from cognitive therapy, such as monitoring negative thoughts, and to use them in an eclectic way. The risk is that the effective components of the approach such as collaboration, structure, focus and homework, may be lost, thus diluting CBT's established effectiveness.

Psychodynamic counselling

Psychodynamic counselling[30] draws from the theoretical traditions of Alderian therapy, Jung's analytical psychology, Freudian psychoanalysis, and Kleinian psychodynamic therapy. Psychodynamic approaches pay particular attention to past experience, particularly adverse relationship experiences during early life, the continuing influence of which may be mediated by unconscious processes. These are seen to influence attachment patterns, psychosocial development, and later psychological functioning. Unconscious processes derived from early experiences contribute to the generation and maintenance of abnormal psychological states. In psychodynamic counselling and psychotherapy, these unconscious processes may be identified through examining

transference and counter-transference in the therapeutic relationship.

The search for the personal meaning of the client's problem or symptoms is central to psychodynamic counselling. The counsellor encourages clients to talk about their difficulties, but also to reflect and gain insight on spontaneous associations and attitudes towards the counsellor as potential sources of information about the presenting problems. Insight alone may be sufficient to enable clients to spontaneously bring about the required changes in their lives. Psychodynamic counselling may also use methods akin to problem-solving and behavioural experiments to facilitate identification and rehearsal of new and more adaptive interpersonal strategies.

Eclectic-integrative approaches

Many practitioners assimilate conceptual and practical way of working that can be attributed to more than one theoretical perspective, formulating the client's difficulties and choosing a mix of methods using more than one theoretical framework. Formally working with a variety of models and methods may be described eclectic or integrative.[31] Such generic therapies often emphasize non-specific factors such as building the therapeutic alliance and engendering hope. Whether this gives the best of each world, or risks the worst of all, is very much open to question, and by nature, eclectic therapy is difficult to standardize for RCTs. The worst kind of *eclecticism* may be an arbitrary pick-and-mix approach, whereby a generically skilled counsellor trains in a variety of approaches and applies these with clients in a way in which he or she deems best. There is little evidence that such an approach is any more effective than the core conditions of counselling allow. Lazarus[32] describes *technical eclecticism*, the drawing of interventions from different sources without necessarily subscribing to their founding discipline. Wherever possible, technically eclectic therapists use treatments based on empirical evidence and client need.

Integration, in contrast, combines identifiable and specific aspects of models in a predetermined way, allowing the evolution of a defined form of therapy such as *cognitive analytic therapy*.[33] *Psychodynamic interpersonal therapy* offers NHS counsellors a point of convergence between predominantly humanistic counselling and more clinically and dynamically orientated approaches often used within psychiatry.[34]

Integration and eclecticism in counselling and therapy will no doubt continue to develop as the nature of clients' problems and the ways of doing therapy evolve. Environmental and technological changes may lead to increasing use of the Internet in counselling and psychotherapy, with face-to-face interactions possible even when client and counsellor are in different locations. It is essential that new counselling approaches are thoroughly supported by empirical evidence so we do not see creeping eclecticism washing out the effectiveness of established methods.

Applications of counselling to specific conditions

Depending on the settings in which they work, counsellors need to be equipped to work with clients with a range of psychological difficulties. For example, the primary care counsellor's caseload is likely to include client difficulties ranging from mild to moderate anxiety or depression to bereavement and relationship problems. Whether or not counselling is an effective intervention for the range of problems seen in these settings has not yet been clearly established. The following looks at counselling for problems where there is good evidence for effectiveness.

Common psychological problems

A Cochrane systematic review[35] compared counselling with normal GP care for people suffering from anxiety, depression, or stress disorders. The authors concluded that overall, significant benefits were seen in mental health improvement from counselling compared with usual GP care, or GP care plus antidepressant treatment, in the short-term (up to 4 months). However these benefits were not maintained in the longer term, over 9 to 12 months. Counselling may be more effective for depression than as a treatment for anxiety. Barrowclough *et al.*[36] compared CBT with supportive counselling (SC) in the treatment of anxiety symptoms in older adults. The CBT group did better than the SC group following treatment, and at follow-up. Overall, cognitively orientated models of counselling and therapy are more effective for depression and anxiety in the long-term than generic counselling. However, counselling may enable people to recover more quickly from depression, and is therefore a valuable and valued intervention.

Counselling for adjustment disorder

Adjustment disorder is defined as a problematic response to a normal stressor, not caused by another mental health problem or bereavement. Such stressors include normal transitions such as leaving home, migration, adverse interpersonal experiences (e.g. relationship breakdown), and unexpected losses such as redundancy. Individual vulnerability can play a part in a person's reaction to life changes, such as previous losses or other adversity, social or cultural isolation, economic deprivation or physical illness.

Counselling is recommended as the first line of treatment for people having difficulty adjusting to life events, illnesses, disabilities or losses, including childbirth and bereavement.[4] The counselling relationship is an important source of security when much has changed in the person's life. The client is helped to identify the stressors, to explore the personal significance of the changes experienced, and to express the emotions generated. It can be necessary to examine unresolved past experiences which may impact on the current adjustment—for example, an individual may not begin to come to terms with redundancy until he recognizes and addresses his unresolved feelings about being abandoned by a parent in childhood. Problem-solving methods are used to identify adaptive goals and ways they may be achieved. The counsellor may encourage the client not to use unhelpful solutions such as denial, excessive use of alcohol, or emotional suppression. The aim is for the client to resolve the crisis themselves.

Relationship problems

Government statistics from both the United States and the United Kingdom show that an ever-growing proportion of marriages fail, with around one in two marriages ending in divorce and an even higher rate in other relationships. Many divorces and relationship problems involve children under the age of 16. A number of specialized services have evolved, including pre-marital counselling, counselling for sexual problems, infertility counselling, bereavement and divorce counselling, and counselling for those involved in

second and subsequent relationships.[37] Telephone counselling and drop in services are frequently used by individuals and couples aiming to clarify the problems and find appropriate help. The kind of issues couples bring to relationship counselling include:

- Communication difficulties
- Conflicts in need between different parties
- Extra-relationship affairs
- Sexual problems
- Conflicts as parents
- Gender role changes
- Violence
- Substance abuse
- Jealousy or possessiveness

Because of the wide range of difficulties, relationship counsellors tend to offer a variety of interventions rather than working within one therapeutic model, and it is unclear whether any one theoretical approach is generally more effective than another.[38] Brief, dynamic work can start the process of internal change, so that the couple is able to work on their own to practise new patterns of relating. Brief, focused work can offer immediate and early solutions to issues which might otherwise threaten the relationship. Longer interventions may be required for couples dealing with major life events and experiences, and offer the opportunity to look at childhood and other roots of persistent or destructive patterns of relating.

The effectiveness of counselling depends a great deal on the willingness of participants to engage in a process which can be painful, challenging often long-established patterns of relating, and one partner in the relationship may be more enthusiastic than the other to promote change. Research shows that many of those who experience relationship counselling understand themselves better, become less emotionally disturbed and understand their partner and relationships better.[39] In some cases, a good outcome is for the relationship to end, in a way which causes least disruption to all parties including children. In the latter case, referral to other agencies may be essential, such as when child protection issues are involved.

Grief counselling

Grief is not a pathological state in itself, and most people emerge from the natural grieving process in a healthy way. Counselling has a role both in facilitating grieving for those who experience difficulties in the process, and in helping those with complex grief reactions.[40] Counselling might involve more than one person in a bereaved family or other grouping, for example the college friends of a student killed in an accident.

In health settings grief counselling is undertaken by trained professionals or volunteers, and in the community by self-help voluntary agencies such as Cruse (in the United Kingdom) or groups attached to hospices.[41] Voluntary agencies are often staffed by people who have themselves experienced bereavement, and group counselling in this context provides a valuable opportunity for acceptance, sharing of experience, and the hope borne out of talking with others who have already come to terms with their loss.

The research on grief counselling[42] shows that professional services and professionally supported voluntary and self-help services can reduce the risk of psychiatric and physical problems following bereavement, and reduce the risk level of 'high-risk' widows to that of a 'low-risk' group. Reid *et al.*[43] found that support by hospice volunteers of high-risk bereaved relatives substantially reduced their levels of anxiety and need for medical care.

Drug and alcohol problems

Drug and alcohol dependence and related problems generate many controversies about their nature and treatment, such as whether people diagnosed as alcoholics can ever return to harm-free, controlled drinking, and the motivation and ability of those addicted to drugs and alcohol to change. Many volunteer agencies offer drug and alcohol treatments, advocating a multifaceted approach, with a combination of methods drawn from motivational interviewing, person-centred and cognitive behavioural therapy known as 'motivational enhancement therapy'. There are several different models of drug and alcohol use, including that of Narcotics Anonymous, similar to Alcoholics Anonymous, who view substance abuse as a pre-existing, biochemical abnormality, necessitating lifelong abstinence. Other views seek to minimize the harm caused by drugs and alcohol, by reducing risks, reducing intake, and possibly changing to another, less harmful substance. A third view sees addiction as a pattern of inappropriate coping: cognitive behavioural principles are used to recognize and deal with situations likely to lead to drug use.

A major trial in the United States randomized 487 patients to one of four, 6-month, treatments. All treatments included group drug counselling following a 12-step model, focusing on achieving abstinence. The group program was offered either alone, or in combination with individual drug counselling, CBT or individual supportive-expressive psychotherapy. Attrition rates in all groups were high, with only 28 per cent completing treatment. All interventions led to a reduction in drug use, but the greatest reduction was for individual plus group counselling.[44]

Counselling for recent and past trauma

Increased public awareness of global trauma arising from natural disasters, war, and terrorism, has led to the development of psychological interventions designed to prevent the onset of post-traumatic stress disorder (PTSD) in those exposed to traumatizing events. However, after many years of early interventions in the form of active single-session 'debriefing' for individuals and groups, there is no evidence for their effectiveness in preventing PTSD. One of the problems in the field is the tendency to 'medicalize' normal distress in traumatic situations, leading to the construction of 'disaster therapists', perhaps with limited understanding of the culture in which the disaster occurred, ready to offer advice and counselling to survivors who may not see themselves as having a mental health problem.[45]

There is also no evidence that non-directive counselling is effective in treating acute stress disorder, which itself constitutes a risk factor for PTSD.[46] Short individualized preventive interventions in the style of 'psychological first aid' may be most effective.

There is no evidence that non-directive or reflective counselling is effective in the treatment of post-traumatic stress disorder. The advocates of counselling for post-traumatic stress disorder describe active-focused methods such as cognitive behaviour therapy,[47]

and using debriefing, to enable people to build a cognitive and emotional account of their experiences. Given that many clients with PTSD are, understandably, mistrustful or avoidant, the methods have to be used within the context of a sound therapeutic relationship, meeting the core counselling conditions.

Postnatal depression

Postnatal depression is often mild and remits spontaneously for many women. However, effective treatment is important because of the potential adverse effect on the child's emotional and cognitive development. Home-based counselling is as effective as antidepressant medication in the treatment of postnatal depression, and is more acceptable to mothers.[48] Counselling can be delivered by health visitors trained in cognitive behavioural counselling methods.[49] The research suggests that depressed mothers benefit from an opportunity to talk about their concerns, not all of which necessarily focus on their baby, with a receptive and non-judgemental professional person. Counselling and other psychological interventions are highly acceptable to mothers with postnatal depression, and preferred over pharmacological treatment. If counselling and other forms of psychotherapy enable women to recover from postnatal depression more rapidly than usual care, this alone may be a valuable service to offer in addition to routine care.

Counselling settings

Counselling takes place in a large number of settings relevant to psychiatry. These include primary care, general medical settings, student counselling services, workplace counselling services, and the voluntary sector. These settings will be described below, aiming to discuss the ways in which counselling may be best adapted to the individual settings, with indications of outcome data on effectiveness.

Counselling in primary care

Primary care is one of the most conspicuous areas of growth in counselling, stimulated by greater demands for alternatives to medication for emotional problems, and by continuing debate about the most effective way of managing mental health problems in primary care. The Layard report on the need for CBT for depression and NICE guidelines stress the importance and effectiveness of psychological interventions including counselling, with a particular focus on CBT. It is likely that the number of counsellors who practice CBT will grow in order to meet demand.

In the United Kingdom, around half of general practices employ a qualified counsellor and the majority meet national criteria for good practice.[50] The development of UK primary care counselling is no doubt part of a wider international trend towards more accessible counselling services at the primary care level.

Counsellors in primary care are a diverse group, in the patients that they see, the counselling models used, and the length of counselling offered. Those identified as primary care counsellors include practice nurses, health visitors, and district nurses trained in counselling skills; clinical and counselling psychologists; community psychiatric nurses, and social workers, and qualified counsellors and psychotherapists. Counselling in primary care is usually provided through one of three main service delivery models[51]:

- Counsellors based in GP practices and provided by a local agency or cooperative
- Managed counselling services provided by the PCT or mental health trust, with counsellors based in GP practices or a central site
- Voluntary agencies (in cases where PCTs have contracts to refer to externally managed services in the voluntary sector)

Counsellors are valued in primary care for a number of reasons,[52] providing time for patients to talk through and reflect on problems, where general practitioners are unable to spend the necessary time on individual patients, as well as a valued alternative or addition to pharmacotherapy. Counselling in primary care also facilitates early identification and intervention for mental health difficulties.

Counsellors in primary care need to be flexible in the way that they work, using different models as appropriate to each client. They also need to be flexible about boundaries and confidentiality, communicating with general practitioners and other health professionals as appropriate. Counselling is generally six to eight 50-min sessions, with a maximum of 20 sessions, and therefore focusing on presenting difficulties rather than long-term issues. The role of the counsellor is varied, and may include offering individual or group counselling, offering advice or training to primary care staff on managing mental health problems, and general consultation.

Despite the growth and popularity of counselling in primary care, it is not clear how effective it is compared to other models such as CBT. Studies have shown mixed findings.[9] Trials comparing counselling for anxiety and depression with usual general practitioner care, CBT, and anti-depressant medication have shown significantly greater clinical effectiveness of counselling compared to usual general practitioner care in the short-term but not in the long-term.[53–55] For people with chronic depression, there were no significant differences between usual care, CBT and short-term psychodynamic counselling,[56] although at 12 months, both psychological therapies were superior to usual care. Counselling in primary care can be cautiously reported as a valuable service, particularly for people with mild to moderate emotional disturbance as well as bereavement and relationship difficulties. Clients improve in the short-term, and the service is appreciated and valued by general practitioners and patients. Primary care services are probably best offered as a range of mental health services, linking closely with community mental health services.

Counselling in general medical settings

Medical patients are understandably at higher risk of psychological difficulties compared to the general population, and many hospital departments and clinics employ counsellors as part of the multidisciplinary team to meet patients' psychosocial needs.[57] Counselling benefits patients in many hospital settings such as gastroenterology, cardiology, obstetrics, and gynaecology, the families of children with medical problems and disabilities and patients with diabetes, renal failure, disfigurement, cancer, head injuries, and chronic conditions such as multiple sclerosis.[58] Counsellors offer what is becoming increasingly limited in medicine—time. A large amount of psychosocial adjustment is needed in most serious illnesses and conditions, and counsellors in health care settings are assisting people cope with potentially life's most challenging moments.

For some conditions, such as HIV/AIDS and genetically transmitted illnesses, counselling forms an important aspect of treatment. From its outset, HIV/AIDS attracted the attention of psychological therapists, since at first there was little else to offer

to help people deal with a strange and potentially fatal illness. A range of professional and voluntary services grew to support those infected. The stages of the illness present different needs, from pre-testing counselling, dealing with the emotional, social, and physical consequences of a positive diagnosis, and managing the adjustment to living with a chronic condition. *Genetic counselling* is specialized branch of counselling practice with increasing applications within psychiatry. In the coming years, the expected identification of susceptibility genes for psychiatric disorders may bring new opportunities and expectations from patients and families for psychiatric genetics.[59]

Counselling is usually offered through referral to a specific hospital department or ward-based counsellor. Many health care workers and hospital chaplains use counselling skills, which are generally regarded as supportive and therapeutic for patients. This does not replace the need for managed counselling services delivered by trained and professional therapists.

Counselling in educational settings

Counselling services in college and university settings cater for students with a wide range of issues. Because of their age and developmental stage, many students have problems adjusting to the new freedoms and demands of college life and also face the developmental challenges of adolescence and young adulthood—conflicts between dependence and independence, psychosexual development, issues to do with self and body image and eating disorders. Other common difficulties include financial, study, and interpersonal problems. Mature students may also contend with the stress of juggling study with children and home-life, and being minorities within a younger peer group. With increasing admissions of overseas students, issues such as identity and loneliness will also arise.

Student counselling services are often arranged so that practical (e.g. financial guidance or careers counselling) and psychological help are offered separately to provide discreet and confidential access. Most clients refer themselves, but may be referred by staff or the student's doctor. Services need to include or work closely with psychiatrists and other mental health professionals in order to meet the needs of students with mental illness. The majority present with less severe emotional or psychological problems, but these may be highly disruptive to their studies and social integration.

Short-term counselling is usually appropriate for students, partly because of their urgency and the structure of the academic year, but also because their natural developmental potential enables most young people quickly to change. This process may be accelerated even more by the intelligence inherent in students, though emotional development can lag behind intellectual development. The task of counselling has been likened to helping the young person back on to the track of normal psychosexual development. More severe derailments, however, may require longer counselling, specialized psychotherapy, or psychiatric treatment.

Counselling in the voluntary sector

Counselling within the voluntary sector has vastly increased, with a growing number of support groups and voluntary organizations offering counselling, mainly on self-referral basis. The most well known in the United Kingdom include Alcoholics Anonymous, Cruse Bereavement Care, Relate for relationship difficulties, and the Samaritans, with equivalent organizations in other countries. Many mental health charities offer support, befriending, and counselling at a 'grassroots' level; some, such as MIND in the United Kingdom, are organized nationally, whereas others operate at a local level. These organizations contribute a great deal to mental health provision, offering the opportunity to talk through and reflect on problems, and in offering support to individuals with more severe mental health problems and their families.

The interface between voluntary and statutory services is varied and at times uncomfortable, with the two sharing different models of care, philosophies, and policies on issues such as confidentiality. There is less research on the effectiveness of counselling provision within the voluntary sector, although one study has shown it to be at least as effective as statutory provision and is often carried out by appropriately trained staff.[60]

Counselling in the workplace

Mental health and employment are known to be significantly related: satisfaction at work is positively correlated with mental health and the unemployed experience higher rates of mental health problems compared to those in employment. The provision of workplace counselling has steadily expanded over the past 20 years, with more than 75 per cent of medium and large organizations in Britain and North America making counselling available to their staff. Counselling may be part of a benefits package, occupational health, human resources or a service brought in to help with specific problems, such as redundancy. Workplace counselling can be viewed as the application of methods of brief psychological interventions that have been shown to be effective in other settings. However, a distinctive strength of seeing a counsellor in the workplace is that the counsellor will be sensitive to the combination of personal and work pressures that the person may present. Workplace counselling is a systemic, as well as individual, intervention in that the organization that pays the counsellor is always present, consciously or unconsciously, influencing the number of sessions and confidentiality boundaries.

Counselling for work-related difficulties is effective in reducing stress-related problems at work and sickness.[61,62] Those who receive counselling are highly satisfied, believing it helps them resolve their problems. Clinically significant improvement in levels of anxiety and depression are reported in 60–75 per cent of clients. Counselling is associated not only with reduction in sickness absence but also improvement in other organizational outcomes such as more positive work attitudes, fewer accidents, and enhanced work performance.

The main provision of counselling at work include employee assistance programmes (EAPs) and specialized staff counselling services.

Employee assistance programmes were first introduced in the United States after the Second World War, to rehabilitate oil-industry employees with alcohol problems. EAPs have become widespread in North America and are increasing in the United Kingdom. They are reported to achieve good results in terms of the percentage of employees who are rehabilitated for work, the reduction in alcohol consumption, improvement of work performance, and cost savings to the company.[63]

EAPs provide a comprehensive confidential counselling service to employees and their families, allowing employee's problems to be identified and resolved at an early stage, and are normally incorporated into the company's benefits package as a form of private emotional health care. EAPs include 24-h access, telephone

counselling, and helplines as well as individual counselling offered at short notice. One of the advantages of counselling organized through EAPs as opposed to in-house staff counselling is improved confidentiality: staff may be reluctant to use counselling services at work if they are not convinced of full confidentiality, and if they fear their career prospects may be adversely affected.

Staff counselling services: many private and public sector organizations now have in-house counselling services.[61] One of the first was set-up by the Post Office in the early 1980s, in recognition of the need to provide emotional and psychological support to employees. Mental health issues, mainly anxiety and depression, formed 46 per cent of the caseload, as well as relationship problems, alcohol problems, bereavement, assault, physical illness or disability, and social problems. Staff counselling schemes are now promoted by many health and education authorities, the Royal College of Nursing, the British Medical Association, and MIND at Work. Problem-solving and cognitive methods of counselling appear to be the most valuable models for workplace settings.[61]

Evaluation of the London Transport Counselling and Trauma Unit showed that the service made huge savings in its first year of operation, in terms of reduced sickness absence and other treatment costs. Research from the United States indicates a return for every dollar invested in Employee Assistance Programmes of between $3 and $7.[61,62] Other benefits may be less quantifiable but nevertheless valuable. For example, a qualitative study[64] of police officers and support staff who had received counselling for work-related difficulties, showed that many described themselves as learning something new and useful about themselves as a result of counselling. For example, an experienced detective stated, *I am 100 per cent better at listening now to a person.*

Telephone and electronically delivered counselling

The telephone is a valuable method of counselling, as shown by organizations such as The Samaritans. Telephone helplines respond to millions of calls each year, and offer a means of talking about feelings and gaining support, information, and advice. The telephone is excellent for crisis intervention and short-term work—one of the key reasons why people phone telephone helplines is because they are in crisis and want to talk anonymously and confidentially.

The telephone and teleconferencing can also be used to conduct some or all of a course of individual or group counselling, or as an adjunct to pharmacological treatment. Telephone advice and structured counselling improved outcome and satisfaction for clients starting antidepressant medication,[65] and also improved adherence to medication.[66]

Electronically delivered counselling: since the mid-1990s, new forms of text technology—the Internet, chatrooms, email, and mobile phone texting—are being developed to deliver counselling and psychotherapy. There is an increasing body of evidence that using text to conduct a therapeutic relationship is not only possible but also in many cases more desirable than face-to-face interaction. Electronically delivered counselling has been addressed by counselling organizations in the United Kingdom, United States and internationally, with published guidelines on issues such as confidentiality and data protection, contracting and informed consent, assessment of suitability of clients, boundaries, and practitioner competence.[67] The International Society for Mental Health Online (http://www.ismho.org/) was formed in 1997 to 'promote the understanding, use and development of online communication, information, and technology for the international mental health community'.

Working electronically has benefits those who cannot access therapy because of, for example, disability or geography. Communicating from a distance can make for a more honest and open relationship, clients diverging information or issues which they would find difficult to discuss face-to-face. Such disinhibition may be empowering—clients having a cathartic experience, so they can then disappear into cyberspace—or potentially hazardous, with traumatic issues remaining unresolved. Although the evaluation of such services is in its infancy, and there are many issues to be addressed, such as practitioner competence and confidentiality, electronic forms of mental health provision are likely to increase. They provide a means of providing help to clients who not only cannot but do not wish to meet with a helper, clients who have been traditionally excluded from mental health provision.

Counselling accreditation, training, and registration

The number of counsellors and psychotherapists in the United Kingdom and other countries is growing rapidly. Professional status and accountability are key concerns about counselling and counsellors, along with other providers of non-medical health care[68]—in the past, anyone could practise as a counsellor, with no standards of training, supervision, or attachment to a regulatory body. As a result, the vulnerable users of such practitioners are at risk. However, there is a strong move within counselling towards clear standards of training and accreditation, to be able to provide an effective and accountable service to the public.

For full *accreditation*, professional counsellors are required to follow, as a minimum, a 3-year full-time training course in the theory and practice of counselling. The standards required are covered by two main organizations in the United Kingdom, with equivalents in other countries:

- The British Association of Counselling and Psychotherapy, which offers accreditation for counsellors from a variety of disciplines, mainly humanistic and psychodynamic.
- The British Psychological Society, which offers accredited chartered status for counselling psychologists, trained to meet standards in applied psychology.

The requirements for accreditation or registration vary between organizations. Common requirements are a minimum of 450 taught contact hours and 450 h of supervised practice, evaluated via case studies and process reports; academic knowledge of counselling theory and research; personal counselling or psychotherapy; and, for the British Psychological Society, research skills and experience. The qualification requires that practitioners follow a code of practice and ethics, stipulating ethical practice, the need for supervision, appropriate confidentiality, and other standards for professional practice.

One of the problems in accreditation is ensuring that counsellors have followed a known and recognized training course. Alongside the expansion of interest in counselling, the number of counselling courses rises each year, ranging from short evening classes in active

listening and short courses in counselling skills for health professionals, to full-time training leading to professional accreditation or chartering. A recent review of UK training found that many psychotherapy and counselling organizations are small, and a significant number of training providers are linked to no external quality assurance systems, 63 per cent having no professional body recognition.[69] There are a large number of titles for both training courses and individual counsellors and psychotherapists, which can only cause confusion to both potential trainees and the public. The review[69] made a number of recommendations to improve the quality of training courses, to lead to greater standardization.

A further change within counselling is the move towards registration. Currently there is no statutory protection for using the terms 'counsellor' or 'psychotherapist', and therefore no means for the prevention of bad practice or abuse. Without registration through a professional body, clients may have no redress for incompetent practice. The National Service Framework for mental health, which sets clear standards for the delivery of effective services, emphasize the importance of 'talking therapies' but stresses the need to register counsellors, psychotherapists, and psychologists. The registering organizations are now moving towards statutory registration and legal protection for the terms counsellor, psychologist, and psychotherapist; and at the time of writing, definitive legislation is likely to be in place by 2008 or so. In the meantime, it is vital that the public and health professionals are aware of the need to seek help only from qualified practitioners of counselling.

Conclusions

For psychiatrists and other mental health care professionals, it can be difficult to make sense of the range of counselling models available, and it is therefore not surprising that potential purchasers and consumers are similarly confused about the nature and advisability of seeking counselling. Counselling has only recently been subject to the rigorous evaluation necessary to meet the standards of evidence-based practice. The most central aspects of counselling, the therapeutic relationship, and qualitative nature of the work, can be difficult to evaluate using established research methodology.

However, counselling is an evolving profession, moving away from, but not forgetting, its roots in Rogerian and psychoanalytic practice, and working hard to meet standards of effectiveness, evaluation, registration, and accountability.

Far more attention is being paid to evaluation, with the development of research paradigms suited to counselling. Systematic evaluation will eventually make it possible to identify, on the basis of clear evidence, the indications for specific models of counselling as well as their limitations. Issues of training, standards, ethics, and accountability are being addressed, to enable counselling to become fully consolidated and integrated within mental health services.

Counselling is a vital part of psychiatry for many reasons. There is a significant gap between the demand for psychological therapy and the available supply. One proposal to overcome this problem is to increase efficiency of provision through the adoption of briefer 'minimal interventions' within stepped care models.[70] Counselling is likely to play an increasingly important part of provision of psychological therapies, particularly for those with mild to moderate mental health problems and social and relationship difficulties. It makes sense for interventions to be offered at an early stage of difficulties: while many problems do resolve on their own, people welcome the support and understanding that counselling can offer, and it is valuable in hastening recovery from emotional distress. Counselling may be appropriate as part of the treatment for people with serious mental illness, offered by psychiatric nurses, social workers, occupational therapists, and workers in the voluntary sector.

Although CBT is being advocated as the treatment of choice for many psychological problems, there is a risk that over-enthusiastic endorsement of the benefits of CBT at the expense of other models may lead to a gap in care: not everyone responds to cognitive approaches, and even when they do, the level of recovery varies. Therefore, the breadth and repertoire of counsellors can add a variety and richness to mental health care provision. *It is this aspect of the human condition, the recognition that we must learn to 'weep for the plague, not just cure it', that is an essential component of meaningful therapy and meaningful relationships. When we experience what seems awful and horrible in our lives, we often take solace in knowing that another person understands, or, at least, is attempting to understand, our pain.*[71]

Acknowledgements

My thanks to Dr Sietske Boeles and Frank Wills for comments on earlier drafts of the chapter.

Further information

Bower, P. and Rowland, N. (2006). Effectiveness and cost effectiveness of counselling in primary care. *Cochrane Database Systematic Review* **3**: Art. No. CD001025.

Feltham, C. and Horton, I. (eds.) (2006). *The Sage handbook of counselling and psychotherapy* (2nd edn). Sage, London.

Nelson-Jones, R. (2006). *Theory and practice of counselling and therapy* (4th edn). Sage, London.

Roth, A. and Fonagy, P. (2005). *What works for whom? A critical review of psychotherapy research* (2nd edn). Guildford Press, New York.

Woolfe, R., Dryden, W., and Strawbridge, S. (eds.) (2003). *Handbook of counselling psychology* (2nd edn). Sage, London.

Resources

The British Association of Counselling and Psychotherapy represents counsellors and psychotherapists in the United Kingdom, with information, training and accreditation: www.bacp.co.uk.
Website for MIND, with information about counselling: www.mind.org.uk.
American Counseling Association: www.counseling.org.
American Mental Health Counselors Association: www.amhca.org.
Canadian Counselling Association: www.ccacc.ca.
Australian Counselling Association: www.theaca.net.au.
European Association for Counselling: www.eacnet.org.

References

1. Chilvers, C., Dewey, M., Fielding, K., *et al.* (2001). Antidepressant drugs and generic counselling for treatment of major depression in primary care: randomised trial with patient preference arms. *British Medical Journal*, **322**, 772–5.

2. Sibbald, B., Addington-Hall, J., Brenneman, D., *et al.* (1996). *The role of counsellors in general practice*. Royal College of General Practitioners, London.
3. Roth, A. and Fonagy, P. (2005). *What works for whom? A critical review of psychotherapy research* (2nd edn). Guilford Press, New York.
4. DOH. (2001). *Treatment choice in psychological therapies and counselling: evidence based clinical practice guidelines*. Department of Health, London.
5. Feltham, C. (2006). What are counselling and psychotherapy? In *The Sage handbook of counselling and psychotherapy* (eds. C. Feltham and I. Horton), pp. 3–10. Sage, London.
6. Burnard, P. (2006). *Counselling skills for health professionals* (4th edn). Nelson Thornes Ltd., Cheltenham.
7. Dryden, W. (2007). *The handbook of individual therapy* (5th edn). Sage, London.
8. Woolfe, R., Dryden, W., and Strawbridge, S. (eds.) (2003). *Handbook of counselling psychology* (2nd edn). Sage, London.
9. Rowland, N. and Goss, S. (eds.) (2000). *Evidence-based counselling and psychological therapies: research and applications*. Routledge, London.
10. McLeod, J. (2002). *Qualitative research in counselling and psychotherapy*. Sage, Thousand Oaks, CA.
11. Ashworth, M., Shepherd, M., Christey, J., *et al.* (2004). A client-generated psychometric instrument: the development of 'PSYCHLOPS'. *Counselling and Psychotherapy Research*, **4**, 27–31.
12. Rogers, C. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting and Clinical Psychology*, **21**, 95–103.
13. Sanders, D. and Wills, F. (2006). *Cognitive therapy: an introduction*. Sage, London.
14. Tolan, J. (2006). *Skills in person-centred counselling and psychotherapy*, p. 18. Sage, London.
15. Krupnick, J., Stosky, S., Simmens, S., *et al.* (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the national institute for mental health treatment of depression collaborative research program. *Journal of Consulting and Clinical Psychology*, **64**, 532–9.
16. Hubble, M., Duncan, B., and Miller, S. (eds.) (1999). *The heart and soul of change: what works in therapy*. American Psychological Association, Washington, DC.
17. Horton, I. (2000). Models of counselling and psychotherapy. In *The Sage handbook of counselling and psychotherapy* (eds. C. Feltham and I. Horton), pp. 234–6. Sage, London.
18. Nelson-Jones, R. (2005). *The theory and practice of counselling psychology* (4th edn). Sage, London.
19. O'Connell, B. (2005). *Solution-focused therapy* (2nd edn). Sage, London.
20. D'Zurilla, T. and Nezu, A. (2006). *Problem-solving therapy: a positive approach to clinical intervention* (3rd edn). Springer Publishing Company, Springer.
21. Mynors-Wallis, L., Gath, D., Day, A., *et al.* (2000). Randomised controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care. *British Medical Journal*, **320**, 26–30.
22. Huibers, M., Beurskens, A., Bleijenberg, G., *et al.* (2003). The effectiveness of psychosocial interventions delivered by general practitioners. *Cochrane Database of Systematic Reviews*, **2**, CD003494.
23. Staring, A., Mulder, C., van-der-Gaag, M., *et al.* (2006). Understanding and improving treatment adherence in patients with psychotic disorders: a review and a proposed intervention. *Current Psychiatry Reviews*, **2**, 487–94.
24. Perls, F., Hefferline, R., and Goodman, D. (1994). *Gestalt therapy: excitement and growth in the human personality*. Souvenir Press, London.
25. Segal, Z., Williams, J., and Teasdale, J. (2002). *Mindfulness-based cognitive therapy for depression: a new approach to preventing relapse*. Guilford, New York.
26. Greenberg, L. (2002). *Emotion focused psychotherapy: coaching clients to work through their feelings*. American Psychological Association, Washington, DC.
27. Novey, T. (2002). Measuring the effectiveness of transactional analysis: an international study. *Transactional Analysis Journal*, **32**, 8–24.
28. Spinelli, E. (2007). *Practising existential psychotherapy: the relational world*. Sage, London.
29. Wills, F. (2006). CBT: can counsellors help fill the gap? *Healthcare Counselling and Psychotherapy*, **6**, 6–9.
30. Jacobs, M. (2006). *The presenting past: the core of psychodynamic counselling and therapy* (3rd edn). Open University Press, Milton Keynes.
31. O'Brien, M. and Houston, G. (2007). *Integrative therapy: a practitioner's guide* (2nd edn). Sage, London.
32. Lazarus, A. (2005). Multimodal therapy. In *Current psychotherapies* (eds. R. Corsini and D. Wedding), pp. 337–71. Thomson Brooks/Cole, Belmont, CA.
33. Ryle, A. and Kerr, I. (2002). *Introducing CAT principles and practice*. Wiley, Chichester.
34. Jenkins, P. (2003). Psychodynamic interpersonal therapy: bridging the gap within the medical model. *Healthcare Counselling and Psychotherapy*, **3**, 15–18.
35. Bower, P., Rowland, N., and Hardy, R. (2003). The clinical effectiveness of counselling in primary care: a systematic review and meta-analysis. *Psychological Medicine*, **33**, 203–15.
36. Barrowclough, C., King, P., Colville, J., *et al.* (2001). A randomized trial of the effectiveness of cognitive-behavioural therapy and supportive counselling for anxiety symptoms in older adults. *Journal of Consulting and Clinical Psychology*, **69**, 756–62.
37. Hill, D. (2006). Relationship problems. In *The Sage handbook of counselling and psychotherapy* (eds. C. Feltham and I. Horton), pp. 453–7. Sage, London.
38. Jacobson N. and Gurman, A. (eds.) (2002). *Clinical handbook of couple therapy*. Guildford, New York.
39. McCarthy, P., Walker, J., and Kain, J. (1998). *Telling it as it is: the client experience of Relate counselling*. Newcastle Centre for Family Studies, Newcastle.
40. Worden, J. (2005). *Grief counselling and grief therapy: a handbook for the mental health practitioner* (3rd edn). Bruner-Routledge, Hove.
41. Parkes, C., Relf, M., and Couldrick, A. (2002). *Counselling in terminal care and bereavement*. BPS Blackwell, Oxford.
42. Parkes, C. (2001). *Bereavement: studies of grief in adult life* (3rd edn). Taylor & Francis, Philadelphia.
43. Reid, D., Field, D., Payne, S., *et al.* (2006). Adult bereavement in five English hospices: types of support. *International Journal of Palliative Nursing*, **12**, 430–7.
44. Crits-Christoph, P., Siqueland, L., McCalmont. E., *et al.* (1999). Psychosocial treatments for cocaine dependence: national institute on drug abuse collaborative cocaine treatment study. *Archives of General Psychiatry*, **56**, 493–502.
45. Summerfield, D. (2006). Survivors of the tsunami: dealing with disaster. *Psychiatry*, **5**, 255–6.
46. Wessely, S. and Deahl, M. (2003). Psychological debriefing is a waste of time. *The British Journal of Psychiatry*, **183**, 12–14.
47. Foa, E., Keane, T., and Friedman, M. (eds.) (2004). *Effective treatments of PTSD*. Guildford, New York.
48. Cooper, P., Murray, L., Wilson, A., *et al.* (2003). Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression. *The British Journal of Psychiatry*, **182**, 412–19.
49. Appleby, L., Hirst, E., Marshall, S. *et al.* (2003). The treatment of postnatal depression by health visitors: impact of brief training on skills and clinical practice. *Journal of Affective Disorders*, **77**, 261–6.

50. Mellor-Clark, J., Simms-Ellis, R., and Burton, M. (2001). *National survey of counsellors working in primary care: evidence for growing professionalization?* Royal College of General Practitioners, London.
51. BACP. (2006). *Shaping effective counselling services in health care: case studies of service delivery and outcomes*. British Association of Counselling and Psychotherapy, Rugby.
52. Keithley, J., Bond, T., and Marsh, G. (eds.) (2002). *Counselling in primary care*. Oxford University Press, Oxford.
53. Bower, P. and Rowland, N. (2006). Effectiveness and cost effectiveness of counselling in primary care. *Cochrane Database of Systematic Reviews*, **3**, CD001025.
54. King, M., Sibbald, B., and Ward, E. (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. *Health Technology Assessment*, **4**, 1–83.
55. Ward, E., King, M., Lloyd, M., *et al.* (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. I: clinical effectiveness. *British Medical Journal*, **321**, 1383–8.
56. Simpson, S., Corney, R., Fitzgerald, P., *et al.* (2000). A randomised controlled trial to evaluate the effectiveness and cost-effectiveness of counselling patients with chronic depression. *Health Technology Assessment*, **4**, 1–83.
57. Thomas, P., Davison, S., Rance, C. (eds.) (2001). *Clinical counselling in medical settings*. Brunner-Routledge, London.
58. Davis, H. and Fallowfield, L. (1996). *Counselling and communication in health care*. John Wiley & Sons, Chichester.
59. Finn, C. and Smoller, J. (2006). Genetic counseling in psychiatry. *Harvard Review of Psychiatry*, **14**, 109–21.
60. Moore, S. (2006). Voluntary sector counselling: has inadequate research resulted in a misunderstood and underutilised resource? *Counselling and Psychotherapy Research*, **6**, 221–6.
61. Greenwood, A. (2006). *Counselling for staff in health service settings*. Royal College of Nursing, London.
62. McLeod, J. and Henderson, M. (2003). Does workplace counselling work? *The British Journal of Psychiatry*, **182**, 103–4.
63. Carroll, M. and Walton, M. (eds.) (1999). *Handbook of counselling in organisations*. Sage, London.
64. Millar, A. (2002). Beyond resolution of presenting issues: clients' experiences of an in-house police counselling service. *Counselling and Psychotherapy Research*, **2**, 159–66.
65. Simon, G., Ludman, E., Tutty, S., *et al.* (2004). Telephone psychotherapy and telephone care management for primary care patients starting antidepressant treatment: a randomized controlled trial. *The Journal of the American Medical Association*, **292**, 935–42.
66. Tutty, S., Simon, G., and Ludman, E. (2000). Telephone counseling as an adjunct to antidepressant treatment in the primary care system. A pilot study. *Effective Clinical Practice*, **3**, 170–8.
67. Goss, S. and Anthony, K. (eds.) (2003). *Technology in counselling and psychotherapy. A practitioner's guide*. Palgrave Macmillan, Basingstoke.
68. DOH. (2006). *The regulation of the non-medical healthcare professions*. Department of Health, London.
69. Aldridge, S. and Pollard, J. (2005). *Interim report to the Department of Health on initial mapping project for psychotherapy and counselling*. British Association for Counselling and Psychotheraphy, Rugby.
70. Bower, P. and Gilbody, S. (2005). Stepped care in psychological therapies: access, effectiveness and efficiency: narrative literature review. *The British Journal of Psychiatry*, **186**, 11–7.
71. Leahy, R. (2001). *Overcoming resistance in cognitive therapy*. Guildford, New York.

6.3.2 Cognitive behaviour therapy

Contents

6.3.2.1 Cognitive behaviour therapy for anxiety disorders

David M. Clark

Introduction

Cognitive behaviour therapy for anxiety disorders is a brief psychological treatment (1 to 16 sessions), based on the cognitive model of emotional disorders. Within this model, it is assumed that it is not events per se, but rather people's expectations and interpretations of events, which are responsible for the production of negative emotions such as anxiety, anger, guilt, or sadness. In anxiety, the important interpretations, or cognitions, concern perceived physical or psychosocial danger. In everyday life, many situations are objectively dangerous. In such situations, individuals' perceptions are often realistic appraisals of the inherent danger. However, Beck[1] argues that in anxiety disorders, patients systematically overestimate the danger inherent in certain situations, bodily sensations, or mental processes. Overestimates of danger can arise from distorted estimates of the likelihood of a feared event, distorted estimates of the severity of the event, and/or distorted estimates of one's coping resources and the availability of rescue factors. Once a stimulus is interpreted as a source of danger, an 'anxiety programme' is activated. This is a pattern of responses that is probably inherited from our evolutionary past and originally served to protect us from harm in objectively dangerous primitive environments (such as attack from a predator). The programme includes changes in autonomic arousal as preparation for flight/fight/fainting and increased scanning of the environment for possible sources of danger. In modern life, there are also situations in which these responses are adaptive (such as getting out of the path of a speeding car). However, when, as in anxiety disorders, the danger is more imagined than real, these anxiety responses are largely inappropriate. Instead of serving a useful function, they contribute to a series of vicious circles that tend to maintain or exacerbate the anxiety disorder.

Two types of vicious circle are common in anxiety disorders. First, the reflexively elicited somatic and cognitive symptoms of anxiety become further sources of perceived danger. For example, blushing can be taken as an indication that one has made a fool of oneself, and this may lead to further embarrassment and

blushing; or a racing heart may be taken as evidence of an impending heart attack and this may produce further anxiety and cardiac symptoms. Second, patients often engage in behavioural and cognitive strategies that are intended to prevent the feared events from occurring. However, because the fears are unrealistic, the main effect of these strategies is to prevent patients from disconfirming their negative beliefs. For example, patients who fear that the unusual and racing thoughts experienced during panic attacks indicate that they are in danger of going mad and often try to control their thoughts and (erroneously) believe that if they had not done so, they would have gone mad.

Within cognitive models of anxiety disorders, at least two different levels of disturbed thinking are distinguished. First, negative automatic thoughts are those thoughts or images that are present in specific situations when an individual is anxious. For example, someone concerned about social evaluation might have the negative thought, 'They think I'm boring', while talking to a group of acquaintances. Second, dysfunctional assumptions are general beliefs, which individuals hold about the world and themselves which are said to make them prone to interpret specific situations in an excessively negative and dysfunctional fashion. For example, a rule involving an extreme equation of self-worth with social approval ('Unless I am liked by everyone, I am worthless') might make an individual particularly likely to interpret silent spells in conversation as an indication that others think one is boring.

Cognitive behaviour therapy attempts to treat anxiety disorders by (a) helping patients identify their negative danger-related thoughts and beliefs, and (b) modifying these cognitions and the behavioural and cognitive processes that normally maintain them. A wide range of procedures are used to achieve these aims, including education, discussion of evidence for and against the beliefs, imagery modification, attentional manipulations, exposure to feared stimuli, and numerous other behavioural assignments. Within sessions there is a strong emphasis on experiential work and on working with high affect. Between sessions, patients follow extensive homework assignments. As in cognitive behaviour therapy for other disorders, the general approach is one of collaborative empiricism in which patient and therapist view the patient's fearful thoughts as hypotheses to be critically examined and tested.

Background

Historical development of cognitive behaviour therapy

Modern cognitive behaviour therapy for anxiety owes its development to pioneering work since the 1950s and 1960s in which the principles of classical conditioning were applied to the understanding and treatment of phobias.[2] It was argued that (a) phobic stimuli are conditioned stimuli that acquired their aversive properties by being paired on one or more occasions with a traumatic event, and (b) avoidance is the main reason why phobias fail to extinguish. This suggestion led naturally to the development of various forms of exposure therapy, in which patients were systematically exposed to phobic stimuli. Initially therapists were concerned that elicitation of strong anxiety responses would be counter-therapeutic so exposure was very gradual, often starting with brief, imaginal presentations, followed by relaxation. Subsequent research showed that such a gentle approach was unnecessary and relatively rapid, *in vivo* exposure became the norm. By the mid-1970s, it was clear that up to 70 per cent of phobics obtained worthwhile improvements from *in vivo* exposure.[3] However, many were less than fully recovered and it was not clear how exposure therapy could be applied to non-phobic anxiety states (such as panic disorder and generalized anxiety disorder). In an attempt to enhance treatment effectiveness further, researchers attempted to identify additional factors that might maintain anxiety. Several cognitive processes outlined below received empirical support. As a consequence, more comprehensive cognitive behavioural treatments that attempt to modify a range of maintaining factors were developed. This chapter describes these treatments.

Cognitive content of anxiety disorders

Although there is no substitute for a careful assessment of each patient's ideation, research shows that most anxiety disorders are characterized by a specific type of fearful ideation and successful therapy generally focuses on such ideation.[4]

(a) Panic disorder

Panic disorder is characterized by a fear of an immediately impending internal disaster (e.g. heart attack, cessation of breathing, mental derangement) and a sense of loss of control over physical and mental functions. Many of panic patients' negative thoughts can be viewed as misinterpretations of normal bodily sensations (such as palpitations or a slight feeling of breathlessness). Indeed, cognitive theorists[5] argue that panic attacks result from a vicious circle in which catastrophic misinterpretations of body sensations lead to an increase in anxiety and associated sensations, which are in turn interpreted as further evidence of impending, internal disasters (e.g. heart attack, fainting, going mad). Panic disorder with agoraphobia is often also accompanied by fear of the interpersonal consequences of attacks (e.g. 'I'll make a fool of myself').

(b) Social phobia

Social phobia is characterized by exaggerated fears of being evaluated, of having one's weaknesses exposed, and of being judged adversely by other people. While in feared social situations, the social phobic continually monitors his or her performance, fears that this performance will be viewed as evidence that he or she is inept, boring, or stupid, and expects that such judgements will have dire long-lasting implications (loss of status or worth and failure to achieve key goals such as friendship, marriage, promotion). Often social phobics have excessively high standards for social performance (e.g. 'My speech must be perfectly fluent', 'I must always appear intelligent and witty'). Typically, social anxiety is triggered when individuals have a strong desire to convey a particular, favourable impression of them and have marked insecurity about their ability to do so.

(c) Generalized anxiety disorder

Generalized anxiety disorder is characterized by excessive worry about a number of life circumstances (e.g. finance, health, work, children, etc.) and the subjective impression that the worry is difficult to control.[6] Beck *et al.*[7] suggested that generalized anxiety disorder patients are anxious about many topics because their beliefs about themselves and the world make them prone to interpret a wide range of situations and circumstances in a threatening fashion. Although their beliefs are quite varied, Beck suggested that they mainly revolve around issues of acceptance, competence, responsibility, and control, as well as the symptoms of anxiety. Borkovec *et al.*[8] have shown that, compared with non-patients,

the worry of general anxiety disorder patients involves less imagery about specific feared outcomes and more verbal rumination in which problems are cast in a more abstract, more difficult to solve, form. Wells[9] has highlighted the importance of positive and negative beliefs about worry (meta-cognition).

(d) Obsessive–compulsive disorder

Obsessive–compulsive disorder is characterized by intrusive and distressing thoughts, impulses, or images about possible harm coming to oneself or others. Thoughts with a similar content to the intrusions of obsessional patients (e.g. a young mother having an intrusive thought about dropping her baby) are common in the general population.[10] For this reason, it has been suggested that the key cognitive abnormality in obsessive–compulsive disorder is not the content of obsessional thoughts, but rather the way the thoughts are interpreted.[11] In particular, it would appear that obsessional patients interpret recurrent obsessional thoughts and impulses as a sign that something terrible will happen, for which they will be responsible. For example, the young mother mentioned above may think that because she had a thought of dropping her baby, she is very likely to do so, despite finding the idea repugnant. In order to prevent the feared consequences of their obsessional thoughts, patients engage in a wide range of 'putting right' acts including (when relevant) washing and checking.

(e) Post-traumatic stress disorder

Surveys[12] indicate that unwanted, intrusive, and distressing memories and the other symptoms of post-traumatic stress disorder (avoidance of reminders and hyperarousal/numbing) are common immediately after traumatic events. Over the next few months many people recover but in a subgroup post-traumatic stress disorder becomes chronic. It is the latter group that normally present for treatment. Research indicates that chronic post-traumatic stress disorder is associated with appraising the traumatic event and/or its sequelae in a manner that would produce a sense of serious current threat to one's view of oneself and/or the world.[13] Examples are given in Table 6.3.2.1.1. There is also evidence that chronic post-traumatic stress disorder tends to be associated with a fragmented memory for the traumatic event and that recovery is associated with developing a more coherent narrative.[14,15]

Why do negative thoughts and beliefs persist?

If the world is not as dangerous as anxiety disorder patients assume, why do they not notice this and correct their thinking? For many patients with chronic anxiety disorders, the persistence of their fears can seem strangely irrational, at least at first glance. Consider, for example, panic disorder patients who think during their panic attacks that they are having a heart attack. Before they come for treatment they may have had several thousand panic attacks, in each one of which they thought they were dying, but they are not dead. Despite what might appear to an outsider as stunning disconfirmation of their belief that a panic attack can kill, their thinking has not changed.

Several factors that appear to prevent patients from changing their negative thinking are outlined below. Such factors are important because reversing them is likely to be a particularly efficient way of treating anxiety disorders.

(a) Avoidance, escape, and safety-seeking behaviours

Early conditioning theorists identified avoidance of, and escape from, feared stimuli as important factors in the maintenance of anxiety disorders. It is easy to see how avoidance of a feared situation (e.g. a supermarket for an agoraphobic) or escape from the situation before a feared event (e.g. a panic attack) occurs could prevent phobics from disconfirming their fears. However, situational avoidance/escape is not so obviously relevant to non-phobic anxiety and some phobics regularly endure feared situations without marked improvement in their fears. Salkovskis[16] introduced the concept of in-situation safety behaviours to deal with this problem. In particular, Salkovskis suggested that while in feared situations most patients engage in a variety of (often subtle) behaviours that are intended to prevent, or minimize, a feared outcome. For example, cardiac concerned panic disorder patients may sit down, rest, and slow down their breathing during attacks and believe, erroneously, that performing these safety behaviours is the reason why they did not die. Experimental studies have confirmed that (a) anxious patients engage in safety behaviours while in feared situations, and (b) dropping these behaviours facilitates fear reduction.[4]

Recent work[17] has highlighted several other important features of safety behaviours. First, although termed 'behaviours', many are internal mental processes. For example, patients with social phobia who are worried that what they say may not make sense and will sound stupid, often report memorizing what they have said and comparing it with what they are about to say, whilst speaking. If everything goes well, patients are likely to think 'It only went well because I did all the memorizing and checking; if I had just been myself people would have realized how stupid I was'. In this way their basic fear persists. Second, it is common for patients to engage in a large number of different safety behaviours while in a feared situation. Table 6.3.2.1.2 illustrates this point by summarizing the safety behaviours used by a patient who had a fear of blushing, especially while talking to men whom she thought other people would think were attractive. Third, safety behaviours can create some of the symptoms that patients fear. For example, responding to a feeling of breathlessness in panic attacks by breathing more quickly and deeply (hyperventilating) can enhance the feeling of being short of breath. Similarly, post-traumatic stress disorder patients who are concerned that unwanted intrusive recollections of the trauma mean they are going mad and often try hard to suppress such recollections. Unfortunately, active suppression increases the probability that the intrusion will occur. Fourth, some safety

Table 6.3.2.1.1 Some examples of idiosyncratic negative appraisals leading to a sense of current threat in post-traumatic stress disorder

What is appraised?	Negative appraisal
Fact that trauma happened	'Nowhere is safe'
One's behaviour/emotions during trauma	'I cannot cope with stress'; 'It was my fault'
Initial post-traumatic stress disorder symptoms	
Irritability, anger outbursts	'My personality has changed for the worse'
Flashbacks, intrusive recollections, and nightmares	'I'm going mad'; 'I'll lose control of my emotions'
Other people's reactions after trauma	
Positive responses	'They think I am too weak to cope on my own'
Negative responses	'Nobody is there for me'; 'I can't rely on other people'

Table 6.3.2.1.2 Safety behaviours associated with a fear of blushing

Feared outcome	Safety behaviour intended to prevent feared outcome
'My face (and neck) will go red'	Keep cool (open windows, drink cold water, avoid coffee, wear thin clothes) Avoid eye contact. If in a meeting, pretend to be writing notes Keep topic of conversation away from 'difficult' issues Tell myself the man is not really attractive. He's no more than a 2 (out of 10)
'If I do blush, people will notice'	Wear clothes (scarf, high collar) that would hide part of the blush Wear make-up to hide the blush Put hands over face. Hide face with long hair Stand in a dark part of the room
'If people notice, they will think badly of me'	Provide an alternative explanation for the red face, e.g. 'it's hot in here'. 'I'm in a terrible rush today', 'I'm recovering from flu', etc.

behaviours can draw other people's attention to problems that patients wish to hide. For example, a secretary who covered her face with her arms whenever she felt she was blushing discovered that colleagues in her office were much more likely to look at her when she did this than when she simply blushed. Finally, some safety behaviours influence other people in a way that tends to maintain the problem. For example, the tendency of social phobics to monitor continually what they have said, and how they think they come across, often makes them appear distant and preoccupied. Other people can interpret this as a sign that the phobic does not like them and, as a consequence, they respond to the phobic in a less warm and friendly fashion.

(b) Attentional deployment

Selective attention plays an important role in maintaining some anxiety disorders. Patients with panic disorder or hypochondriasis fear certain bodily sensations and symptoms, believing they indicate the presence of a serious physical disorder (heart attack, cardiac disease, cancer, etc.). Such patients have often had several medical investigations that indicate they do not have the physical illness(es) they fear, but they are not convinced. One reason appears to be that their fears lead them to focus attention on relevant parts of their bodies and, as a consequence of this attentional deployment, they become aware of benign bodily sensations that other people do not notice.[5] The presence of such sensations is then taken by the patient as evidence that a serious physical illness has been missed. (Hypochondriasis is classified as a somatoform disorder in DSM-IV[6] and as a somatization disorder in ICD-10.[18] However, it has many features in common with anxiety disorders and can be conceptualized as such for the purposes of psychological treatment).

Social phobia appears to be associated with two attentional biases. First, when in feared social situations, patients with social phobia report becoming highly self-focused, constantly monitoring how they think and feel they are coming across, and paying less attention to other people.[17] Reduced processing of other people means that social phobics have less chance to observe other people's responses in detail and, therefore, are unlikely to collect from other people's reactions information that would help them to see that they generally come across more positively than they think. Second, there is some evidence that when social phobics do focus on other people, they are particularly good at detecting negative reactions[19] and are poor at detecting positive reactions.

(c) Spontaneously occurring images

Spontaneously occurring images are common in anxiety disorders and also appear to play a role in maintenance. Patients with social phobia often report 'observer-perspective' images in which they see themselves as if viewed from outside.[20] Unfortunately, in their images they do not see what a true observer would see, but rather their fears visualized. For example, a teacher who was anxious about talking with colleagues in coffee breaks noticed that before speaking she felt tense around her lips. The tension would trigger an image in which she saw herself with a twisted and contorted mouth, looking like 'the village idiot'. At that moment, she was convinced everyone else thought she was stupid. Negative images are also used as information in other anxiety disorders. For example, obsessional patients who have images of committing a repugnant act (e.g. stabbing one's child) take the occurrence of the image as evidence that they are in danger of performing the act. Similarly, patients with post-traumatic stress disorder report that flashbacks increase the perceived likelihood of a future trauma.

(d) Emotional reasoning

A further source of misleading information that can enhance patients' perception of danger is anxiety itself.[21] For example, social phobics often think they look as anxious as they feel, but in general this is not the case.[22] Similarly, generalized anxiety disorder patients often take feeling on edge as a sign that something bad is about to happen.

(e) Memory processes

Some anxiety disorders are associated with a tendency for the selective recall of information that would appear to confirm the patient's worst fears. For example, high socially anxious individuals selectively recall negative information about the way they think they have appeared to others in the past when anticipating a stressful social interaction.[22] Similarly, patients with hypochondriasis selectively recall illness-related information. In post-traumatic stress disorder, a failure to elaborate memories at the time of the trauma and enhanced associative learning appear to play a key role in maintaining the re-experiencing symptoms.[13]

(f) Rumination

Anxious patients often spend protracted periods of time ruminating about negative things that could happen in the future and about how bad they would be. They may also ruminate about things that they feel have gone wrong in the past. Studies by Davey and Matchett[23] indicate that such rumination can enhance fear. There are several ways in which rumination might operate. First, thinking about an event may directly increase its subjective probability. Second, selectively focusing on past negative events, feelings, and impressions may further enhance the perceived likelihood of future danger. Third, rumination is rarely focused on constructively processing perceived threats, but instead often seems to elaborate the

threats or make them more abstract and hence difficult to deal with. For example, patients with post-traumatic stress disorder often ask themselves 'Could I have done something different?' during their traumatic event without thinking through in detail what their alternative options might have been, and how feasible they would have been at the time.

Treatment

Assessment interview

Table 6.3.2.1.3 summarizes the main topics covered in the assessment interview. The aims of the interview are as follows: (a) to obtain a detailed description of the patient's fears and behaviour; (b) to identify maintaining factors; (c) to normalize the problem; (d) to develop a model of the problem that can be used to guide treatment.

The interview would start by asking the patient to provide a brief description of the main presenting problem(s). For example, intense anxiety attacks, anxious apprehension, and avoidance of places where the attacks seem particularly likely or would be embarrassing. The interviewer then obtains a detailed description of a recent occasion when the problem occurred or was at its most marked. This would include the situation ('Where were you?', 'What were you doing?'), bodily reactions ('What did you notice in your body?', 'What sensations did you experience?'), thoughts ('At the moment you were feeling particularly anxious, what went through your mind? What was the worst that you thought might happen? Did you have an image/mental picture of that? How do you think you looked?'), behaviour ('What did you do?'), and the behaviour of others ('How did X react?', 'What did X say/do?'). Having obtained a detailed description of a recent occasion, the interviewer should check whether the occasion was typical. If not, further descriptions of other recent occasions should be elicited to provide a complete picture.

Table 6.3.2.1.3 Summary of topics to be covered in assessment interview

Brief description of presenting problem(s)
For each problem
Detailed description of a recent occasion when problem occurred/was at its most marked
Situation
Bodily reaction
Cognitions
Behaviour
List of situations when the problem is most likely to occur/be most severe
Modulators (things making it better or worse)
Possible maintaining factors
Avoidance of situations/activities
Safety behaviours
Attentional deployment
Faulty beliefs
Attitudes and behaviour of others
Medication
Beliefs about cause of the problem
Previous treatment (types, whether successful)
Onset and course
Personal strengths and assets
Social and financial circumstances

Next a list of situations in which the problem is most likely to occur or is most severe is elicited ('Are there any situations in which you are particularly likely to have a panic attack?'), together with information about modulators ('Are there any things that you notice make the symptoms stronger/more likely to occur?', 'Are there any things that you've noticed make the symptoms less likely/less severe/more controllable?').

Possible maintaining factors should be identified, including the following:

- avoidance of situations or activities ('What situations/activities do you avoid because of your fears?')
- safety behaviours ('When you are afraid that X might happen, is there anything you do to try to stop it happening?')
- attentional deployment ('What happens to your attention when you are worried about X? Do you focus more on your body? Do you become self-conscious?')
- faulty beliefs (e.g. an obsessive–compulsive disorder patient, believing that thinking something can make it happen)
- attitudes and behaviour of significant others ('What does Y think about the problem?'; 'What does Y do when you are particularly anxious?')
- current medication

There are several ways in which excessive use of both prescribed and non-prescribed medications can maintain anxiety disorders. For example, painkillers and tranquillizers can cause derealization and sleep disturbance respectively, and drinking before social occasions prevents disconfirmation of one's social fears.

It is also important to assess patients' beliefs about the cause of their problems as some beliefs may make it difficult for patients to engage in therapy. For example, patients with post-traumatic stress disorder who think the best way of dealing with a painful memory is to push it out of their mind are unlikely to engage in imaginal reliving of the event until this belief is dealt with.

Finally, a brief description of the onset and subsequent course of the problem should be obtained. This description should particularly focus on factors, which may have been responsible for initial onset and for fluctuations in the course of the symptoms and is primarily used to make the development of the problem seem understandable to the patient.

It is not always possible to obtain all the information needed for a cognitive behavioural formulation in an assessment interview. Sometimes it is necessary to follow-up the interview with homework assignments in which the patient collects more information to clarify the formulation. For example, a hypochondriacal patient who was concerned that palpitations meant that she had cardiac disease was asked to record what she did each hour and how many palpitations she experienced. To her surprise, palpitations were not associated with exercise, as she expected, but rather were most common when she was sitting quietly, reading, watching television, or studying. This realization helped convince her that her problem may be disease preoccupation rather than a faulty heart.

Developing an idiosyncratic model of the patient's problem

Assessment ends with the development of an idiosyncratic version of the cognitive model. In particular, therapists aim to show patients

how the specific triggers for their anxiety produce negative automatic thoughts relating to feared outcomes and how these are maintained by safety behaviours and other maintenance processes. The model is usually drawn on a whiteboard, so that patient and therapist can look at it and discuss it together. Figure 6.3.2.1.1 shows an example for a panic disorder patient. His panic attack started with a twinge in his chest muscles, and he then had the thought, 'There is something wrong with my chest area, maybe I am having a heart attack'. This interpretation made him start to feel anxious, his chest muscles tightened up more, he started to feel dizzy, his heart raced more, and he then thought, 'I'm dying, I'm having a heart attack', and also, interestingly, 'If I don't die, people will notice I'm anxious and think it is odd'. He then engaged in a series of safety behaviours to try to prevent himself from dying. He thought he had read somewhere that paracetamol (aminacetophen) is good for people with heart problems and so he took a paracetamol. This is incorrect information, but the key point is that he believed it. He also sat down and rested, took the strain off his heart, and took deep breaths, trying to slow down his heart rate. He believed that the main reason he had not died was that he had engaged in the safety behaviours. The reader will also notice that some of the safety behaviours (taking deep breaths and monitoring the heart) will also have augmented his feared symptoms.

Figure 6.3.2.1.2 shows a further example with a social phobic patient. The patient's main fear was that other people would think she was stupid and boring. The situation used to develop the model was a recent coffee break at work during which the patient had difficulty joining a conversation with colleagues. When attempting to join the conversation she had the thought, 'I'll sound stupid and everyone will think I am dumb'. In order to prevent herself from sounding stupid, she engaged in an extensive set of safety behaviours which (a) prevented her from discovering that her spontaneous thoughts are interesting to other people, (b) made her appear preoccupied and uninterested in her colleagues, and (c) made her excessively self-conscious. While self-conscious, she became particularly aware of anxiety symptoms (sweaty palms, stiff muscles around her mouth) that she thought other people might see, and indeed, had an image of herself in which she looked very strange, with a twisted and rigid mouth and appeared stupid.

Normally idiosyncratic models of the form illustrated in Figs 6.3.2.1.1 and 6.3.2.1.2 will be developed at the end of the first interview, and certainly not later than the second session. Such models are used as blueprints to help therapist and patient organize and develop the rest of therapy.

Monitoring progress

Once treatment has started, it is important to monitor progress continually in order to decide whether a particular treatment procedure is working or whether the case needs reformulating and new treatment procedures need to be implemented. Usually patients are asked to complete a small number of self-report questionnaires before each therapy session. Typically, these include frequency and severity ratings for the main anxiety problems (often using simple 0–8 Likert-type scales), a measure of negative thoughts, and general measures of anxiety and depression (such as the Beck Anxiety Inventory[24] and the Beck Depression Inventory[25]). Table 6.3.2.1.4 summarizes some of the most commonly used weekly measures. In some instances these are supplemented by more individualized diaries and ratings. More global standardized measures of symptom severity are also often administered at the beginning, middle, and end of therapy in order to provide normative data (see Table 6.3.2.1.4).

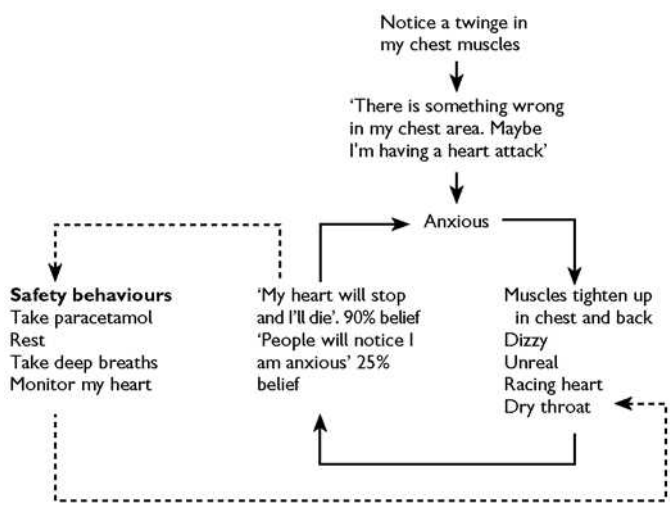


Fig. 6.3.2.1.1 A cognitive model of a patient's panic attacks. (Reproduced from Clark, D.M., Panic disorder: from theory to therapy. In *Frontiers of cognitive therapy* (ed. P.M. Salkovskis), pp. 318–344, Copyright 1996, Guildford Press, New York.)

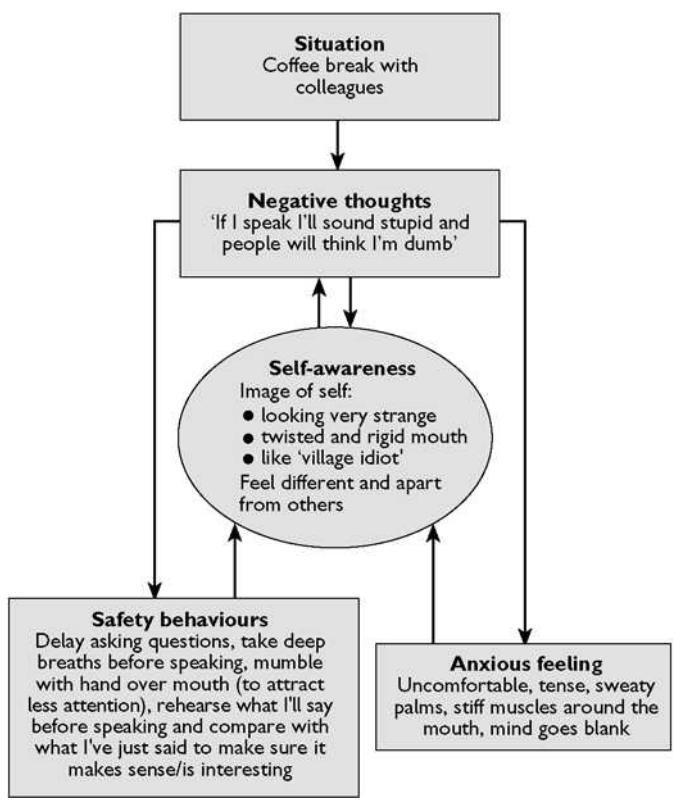


Fig. 6.3.2.1.2 A cognitive model for a patient with social phobia.

Treatment procedures

A wide range of procedures can be used to modify patients' negative beliefs and linked maintenance processes. For clarity the procedures are described separately. However, in practice the techniques are closely interwoven. Within a given session, therapists will usually use a mixture of discussion and experiential techniques to help patients to challenge convincingly their negative beliefs. As with cognitive behaviour therapy for other disorders, patients are given extensive homework assignments and it is assumed that a sizeable amount of therapeutic change is the result of homework assignments.

(a) Identifying patients' evidence for their negative beliefs

Anxiety disorder patients usually have reasons for believing that the things they fear are dangerous, however strange their fears may seem. The therapist, therefore, tries to 'get inside the patient's head' and see what the evidence is. Often the evidence is an event or piece of information that the patient has misinterpreted. Identifying and correcting such misinterpretations can be helpful. For example, a panic disorder patient believed that experiencing high anxiety could kill her. When asked by the therapist what her evidence was, she explained that she had seen it happen. Further enquiry revealed she had entered Dresden the day after the fire bombing of that city by the allies during the Second World War and had helped search for survivors. When opening up cellars below demolished houses, she repeatedly observed that the occupants were either dead or behaved in a dazed confused manner, even though the fire had not entered their cellars. She concluded that fear had killed the occupants or sent them mad. However, further questioning from the therapist revealed that the cellar occupants all had bright cherry-red lips. This allowed the therapist to explain that they were suffering from carbon monoxide poisoning, not the effects of intense fear. This correction considerably reduced the patient's fear of anxiety.

(b) Education

Education about the symptoms of anxiety is often helpful, especially if it directly targets patients' idiosyncratic fears and concerns. For example, post-traumatic stress disorder patients often think their flashbacks and emotional outbursts mean they are going mad or have permanently changed for the worse. In such cases, detailed assessment of the patient's post-traumatic stress disorder symptoms and explanation that each are common reactions to a trauma can greatly help. Similarly, panic disorder patients with cardiac concerns often cite left-sided chest pain as evidence for their belief that they have a cardiac disorder. In such cases discussion of Fig. 6.3.2.1.3 (from a study of chest pain in patients referred to a cardiac clinic[26]) is useful. In particular, the patient discovers that left-sided chest pain is more characteristic of non-cardiac chest pain than of either confirmed angina or myocardial infarction. Further questioning helps patients to see that the association between left-sided pain and attacks is probably a consequence of their fears. That is to say, they can experience pain on either side of the chest but only panic when it is on the left side. Finally, patients with obsessive–compulsive disorder who are perturbed by the apparently repulsive and unusual nature of their intrusive thoughts often benefit from reviewing Rachman and De Silva's classic

Table 6.3.2.1.4 Commonly used measures for monitoring progress

Anxiety disorder	Measure		
	Symptoms	**Thoughts**	**Global severity**
Panic disorder	Panic Rating Scale[63] Panic Diary[63] BAI[33] BDI[2]	Agoraphobic Cognitions Questionnaire[65]	Fear Questionnaire[64] Mobility Inventory[66]
Social phobia	Social Summary Scales (Table 4) BAI[33] BDI[2]	Social Cognitions Questionnaire[53]	Liebowitz Social Anxiety Scale[67] Social Performance Scale[68] Social Interaction Anxiety Scale'[68]' Social Phobia and Anxiety Inventory[69]
Generalized anxiety disorder	BAI[33] BDI[2]	Worry Domains Questionnaire[70] Thought Control Questionnaire[72]	Penn State Worry Questionnaire[71] Spielberger State Trait Inventory[73]
Obsessive- compulsive disorder	BAI[33] BDI[2]	Responsibility Interpretations Questionnaire[74]	Padua Inventory[75] Yale–Brown Obsessive Compulsive Scale[76]
Post-traumatic stress disorder	Post-traumatic Diagnosis Scale[77,78] BAI[33] BDI[2]	Post-traumatic Cognitions Inventory[79] Personal Beliefs and Reactions Scale[81]	Impact of Events Scale[80] Post-traumatic Diagnosis Scale[78]

Owing to their length, the Post-traumatic Cognitions Inventory and the Personal Beliefs and Reactions Scale are not suitable for weekly administration.

BAI, Beck Anxiety Inventory; BDI. Beck Depression Inventory.

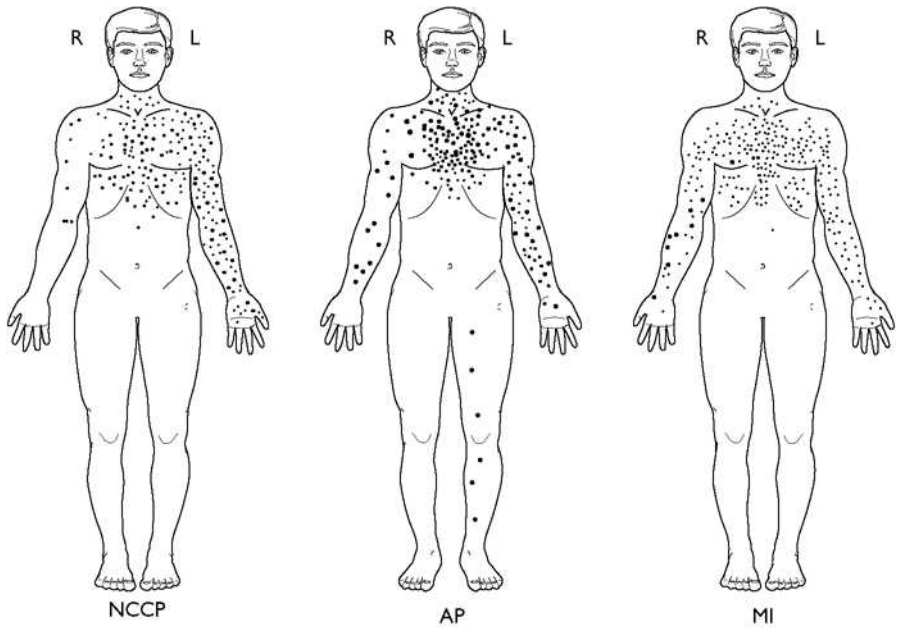


Fig. 6.3.2.1.3 Distribution of chest pain in patients referred to a cardiac clinic and subsequently diagnosed as non-cardiac chest pain (NCCP), angina pectoris (AP), or myocardial infarction (MI). (Reproduced from Beunderman, R. *et al.* (1998), Differentiation in prodromal and acute symptoms of patients with cardiac and non-cardiac chest pain, In *Advances in theory and practice in behaviour therapy* (ed. P.M.G. Emmelkamp, *et al.*), Copyright 1998, Swets and Seitlinger, Taylor and Francis Group, an informa business.)

paper[10] which demonstrated that thoughts with identical content to obsessional intrusions are common in the general population.

(c) Identifying observations that contradict patients' negative beliefs

As anxiety disorder patients' beliefs about the dangerousness of feared stimuli are generally mistaken, patients have often experienced a number of events that contradict their beliefs before they come into therapy. Therapists can make considerable progress, even in an assessment interview, by spotting these events and helping patients understand their significance. For example, panic disorder patients who are worried that their symptoms mean they are about to have a heart attack, often report that in some attacks something unexpected happened to distract them (e.g. a telephone call) and then their symptoms went away. Therapists could then pause and help the patient understand what this means, perhaps asking, 'Would a cardiologist prescribe telephone calls as a treatment for a heart attack?' The patient would probably answer 'No', to which the therapist might reply, 'If telephone calls would not stop a heart attack, how might they work? If the problem was the negative thought, could they help (by distracting one from the thoughts)?'

(d) Imagery modification

Images play an important role in many anxiety disorders. Most images represent feared catastrophes and can be treated as predictions to be tested (see behavioural experiments below). However, when the images are stereotyped and repetitive it is often also necessary to work directly with the images and to restructure them explicitly.

The problem with anxiety-related images is that they seem very realistic at the time they occur and, as a consequence, greatly enhance fear. A common restructuring technique involves discussing with the patient whether the image is realistic. Once it is intellectually agreed that the image is an exaggeration, patients are asked to recreate intentionally the negative image and to hold it in mind until they start to feel anxious. They are then asked to transform it into a more realistic image, or an image, which convincingly indicates that the original image was unrealistic. A common observation is that patients' spontaneous images generally stop at the worst moment. For example, agoraphobic patients who fear fainting in a supermarket might see themselves collapsed on the floor, but not see themselves getting up, recovering, and going home. A useful transformation in such cases is to 'finish out' the image by asking patients to run it on until they see the positive resolution. Of course, sometimes simply running on an image does not produce a positive resolution. For example, a patient who feared she would go mad frequently experienced an image of two men in white coats entering her house to take her away to a locked ward. In the image, the men were extremely powerful and she felt powerless. Transformation, following suggestions from her, involved shrinking the men and then turning them into ridiculous looking (and hence non-threatening) white poodles.

An interesting observation about spontaneous imagery is that it often fails to incorporate positive information that would seriously undermine the impact of the image, even when the patient has such information. For example, a mother whose children died in a house fire, repeatedly experienced intrusive flashbacks in which she saw the house going up in flames and smelled burning flesh, despite having seen her children in the mortuary, knowing that they had not been burnt, but instead were rapidly overcome by fumes.

For imagery restructuring to be effective it is important that it is not done as a cold, intellectual exercise, but instead includes eliciting the affect normally associated with the image. Transformation may have to be done in several steps. It is often best to start with the

most threatening aspect of the image. Possible alternative images should be generated by patients, rather than simply imposed by the therapist.

(e) Cognitive restructuring

All the above techniques are examples of cognitive restructuring in which the therapist provides information and asks a series of questions to help the patients challenge their fearful thoughts and images. A list of some of the questions that can be particularly useful for helping anxiety disorder patients challenge their negative thoughts is given in Table 6.3.2.1.5. Further useful questioning techniques can be found in Chapter 6.3.2.3.

It is sometimes helpful to use graphical methods for discussing alternatives to negative thoughts. In situations where there are several non-threatening alternative explanations for a feared event, pie charts are particularly useful. When constructing a pie chart the therapist draws a circle which is meant to represent all the possible causes of a particular event and asks the patient to list all the possible non-catastrophic causes of the event and allocate a section of the circle to each cause. At the end of the exercise, there is often very little of the circle left for the patient's negative explanations. Figure 6.3.2.1.4 illustrates the use of a pie chart to challenge a generalized anxiety disorder patient's belief that he would be 100 per cent responsible for people not enjoying themselves at his dinner parties. The belief was preventing him from making new social contacts after a painful divorce. Pie charts are particularly helpful for dealing with distorted beliefs about responsibility and hypochondriacal concerns (e.g. 'Headaches mean I have a brain tumour').

When considering the worst that could possibly happen in a feared situation patients frequently ignore the fact that there are many intermediate events, each with a probability of less than 1, which have to occur for the catastrophe to be realized. The inverted pyramid can be a good way of representing this. Figure 6.3.2.1.5 shows an example with a patient who was afraid of blushing. His worst fear was that other people would think they were greatly superior to him if he blushed. Whenever he felt his face becoming hot, he was convinced other people were thinking they are superior to him and gloating. However, careful discussion helped him to see that there were many intermediate steps between him feeling hot and the feared outcome. Once the conditional probabilities were taken into account, there was only a minute chance that his worst fear would be correct.

Table 6.3.2.1.5 Useful questions for challenging anxiety-related thoughts

What is the evidence for this thought?
Is there any alternative way of looking at the situation?
Is there an alternative explanation?
How would someone else think about the situation?
Are you focusing on how you felt, rather than on what actually happened?
Are you setting yourself an unrealistic or unobtainable standard?
Are you forgetting relevant facts or overfocusing on irrelevant facts?
Are you thinking in all-or-nothing terms?
Are you overestimating how responsible you are for the way things work out?
What if the worst happens? What would be so bad about that? How could you cope?
How will things be x months/years afterwards?
Are you overestimating how likely the event is?
Are you underestimating what you can do to deal with the problem/situation?

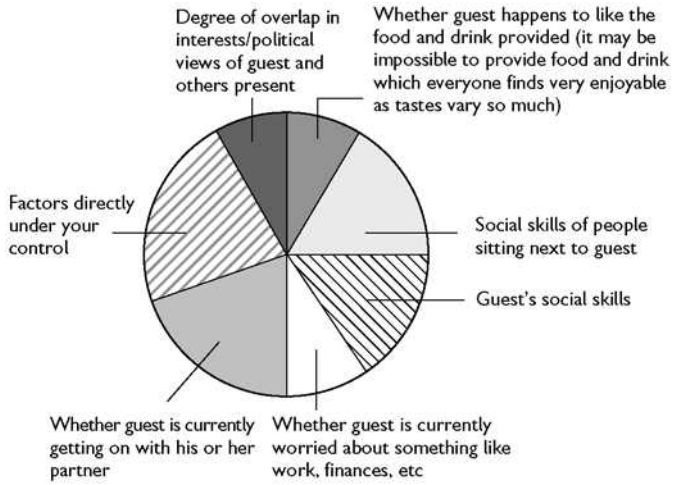


Fig. 6.3.2.1.4 A pie chart representing factors that might contribute to guests enjoying themselves at a dinner parties to challenge a generalized anxiety disorder patient's belief that he would be 100 per cent responsible for people not enjoying themselves. (Reproduced from Clark, D.M., Anxiety states: panic and generalized anxiety, In *Cognitive therapy for psychiatric problems: a practical guide* (ed. K. Hawton, *et al.*), pp. 52–96. Copyright 1989 with permission from Oxford University Press.)

It is important to remember that anxiety results from overestimating the cost of feared events as well as their probability. Discussions aimed at modifying perceived cost are often helpful. This can be true even in cases where it might seem obvious that the feared event is objectively costly. For example, in hypochondriacal patients who are worried about dying, therapists may be tempted to focus exclusively on whether or not the patients are likely to die from the symptoms they are concerned about. Accepting that dying is a bad thing, the therapist may not be inclined to ask, 'What would be so bad about dying?' However, Wells and Hackmann[27] found that many hypochondriacal patients have distorted beliefs and images about death and the process of dying. For example, they think that when they die they will remain conscious and will continue to experience all the pain they had up to that point. Such people can benefit greatly from discussion of their beliefs about the cost of dying.

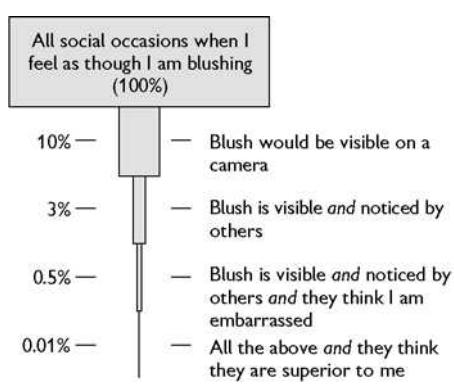


Fig. 6.3.2.1.5 An inverted pyramid representing conditional probabilities between 'feeling hot' and 'others thinking they are superior to you' constructed for a patient who was afraid of blushing.

(f) *In vivo* exposure to feared situations, activities, and sensations

Systematic exposure to feared and avoided situations has a long history in cognitive behaviour therapy and is one of the most effective ways of helping patients to discover that the things they are afraid of will not happen or are more manageable than they anticipate. Initially, exposure was often conducted in imagination but it is now known that *in vivo* exposure is a more effective way of dealing with situational fear.[3] During the 1970s and 1980s the dominant framework for exposure was habituation. It was assumed that repeated prolonged exposure was required to achieve fear reduction. More recent cognitive formulations have suggested that exposure is likely to be optimally effective when set-up in a way that maximizes the extent to which patients are able to disconfirm their fears, and considerable attention is now devoted to setting up exposure assignments in a way that will maximize cognitive change. Before entering a feared situation, patients are asked to specify what is the worst they think could happen, how likely they think it is, and what they would normally do to prevent the feared catastrophe (safety behaviours). They are then asked to enter the feared situation while dropping their safety behaviours and to observe carefully whether the feared outcome occurs. Afterwards, discussion focuses on whether the feared catastrophe occurred. If it did not, how does the patient explain its non-occurrence? Was it because the patient now thinks the feared outcome is unrealistic or does the patient think it was because of 'luck' or the continued use of safety behaviours? In the latter two instances, further exposure assignments with further encouragement to drop safety behaviours are required.

In addition to avoiding feared situations, anxiety disorder patients can also avoid feared sensations. Such avoidance is particularly prominent in panic disorder. For example, because of their fears about the meaning of increases in heart rate, dizziness, sweating, and other autonomic cues, panic disorder patients often avoid exercise. Increasing exercise can be an excellent way of helping them to challenge their negative beliefs, as can other ways of inducing bodily sensations such as ingesting caffeine, and hyperventilating. In each instance, the key point is to help patients discover that they can experience intense physical sensations without dying, losing control, or experiencing some other catastrophe.

Table 6.3.2.1.6 shows a record sheet that can be useful for planning and summarizing the results of exposure assignments, with illustrations from patients with social phobia and agoraphobia. Because of the intensity of patients' fears, and their tendency to attribute good outcomes to 'luck', it is often necessary to move up a hierarchy of feared situations and to consolidate successes by repetition.

In obsessive–compulsive disorder, the compulsive rituals act as safety behaviours and it is necessary to ensure that patients refrain from engaging in rituals (which are often also termed 'putting right' acts) during exposure assignments. This procedure is called 'exposure and response prevention'. For example, obsessional washers would be asked to 'contaminate' themselves by touching feared objects and then not put things right by washing. Similarly, obsessional checkers may be asked to expose themselves to activities that would normally provoke their checking (e.g. turning on the gas cooker) and then refrain from checking more than would be normal. In both instances, patients usually find that although exposure initially provokes considerable distress, the distress systematically declines during prolonged response prevention.[28]

Unlike most phobic fears, the fears of obsessive–compulsive disorder patients (e.g. developing a fatal disease from touching an object that is believed to be contaminated) often cannot be disconfirmed during a single or indeed multiple, exposure assignments. Discussing this issue, Salkovskis[11] has suggested that exposure and response prevention may work by providing patients with a different understanding of their problems. In particular, the decline

Table 6.3.2.1.6 Record sheet for noting behavioural experiments

	Situation	Prediction	Experiment	Outcome	What I learned
		(What exactly did you think would happen? How would you know?) (Rate belief 0–100%)	(What did you do to test the prediction?)	(What actually happened? Was the prediction correct?)	(1) Balanced view (rate belief 0–100%)? (2) How likely is what you predicted to happen in future (Rate belief 0–100%)
Social phobic patient	Coffee break; sitting with other teachers; trying to join in the conversation	If I just say things as they come into my mind, they'll think I'm stupid (50%)	Say whatever comes into my mind *and* watch them like a hawk; don't focus on myself; this only gives me misleading information (such as images of myself as the 'village idiot'), and means I can't see them	I did it and I watched the others; one of them showed interest and we talked; she seemed to quite enjoy it	I am probably more acceptable than I think (70%)
Agoraphobic patient	Shopping in a supermarket	I will feel dizzy and have a panic attack (90%). Unless I grip the trolley tightly or sit down at that moment, I collapse (80%)	Go into the supermarket. When I start to feel dizzy, remind myself it is just anxiety, my heart rate is up and I can't faint. Then move away from the trolley and stand unsupported	I felt dizzy but didn't faint, even though I didn't sit down or hold on to the trolley	Feeling dizzy in anxiety attacks will not make me faint (60%)

in distress during exposure and response prevention helps the patient to discover that they are suffering from a worry problem, rather than being in objective danger.

(g) Imaginal exposure in post-traumatic stress disorder

Although imaginal exposure is rarely used in most anxiety disorders, it plays an important role in the treatment of post-traumatic stress disorder. It is known[29] that avoidance of thinking about the traumatic event is an important predictor of persistent post-traumatic stress disorder. In the light of this finding, clinicians have attempted to treat post-traumatic stress disorder by repeated, imaginal reliving of the traumatic event, and controlled trials[30] have shown that this technique is effective. At this stage it is not known why reliving works. One suggestion is that the intrusive symptoms of post-traumatic stress disorder are the result of a fragmented and disorganized memory for the trauma that is poorly integrated with other autobiographical information. Reliving might, therefore, facilitate the production of a more organized narrative account of the event that can be placed in the broader context of the individual's life.[13,31] Two types of imaginal reliving have been used in controlled trials: writing out details of the event and reliving the event in imagery. In either case, it seems important to focus not only on what happened, but also on patients' feelings and thoughts, both at the time and now, looking back at the event. Problematic idiosyncratic meanings that can be addressed with cognitive restructuring are often identified during reliving exercises.

(h) Behavioural experiments

Behavioural experiments play a central role in the treatment of anxiety disorders. In a behavioural experiment, therapist and patient plan and implement a behavioural assignment that will provide a test of a key belief. The *in vivo* exposure assignments outlined above are examples of behavioural experiments. Several further examples are given to illustrate the technique.

Patients with post-traumatic stress disorder often think their intrusive recollections mean they are going mad or losing control in some way, and as a consequence, try to push the intrusions out of their mind. If this problem is identified during the first session of therapy, therapists often conduct an experiment to illustrate the undesired consequences of thought suppression. For example, the therapist might say to the patient, 'It doesn't matter what you think about in the next few minutes as long as you don't think about one particular thing. The thing is a fluorescent green rabbit eating my hair!' Most patients find they immediately get an image of the rabbit and have difficulty getting rid of it. Discussion then helps them to see that an increase in the frequency of target thoughts is a normal consequence of thought suppression. This result can then be used to set-up a homework assignment in which the patient is asked to collect data to test the idea that thought suppression may be enhancing intrusions. The experiment involves not trying to push the intrusions out of one's mind, but instead just letting them come and go, watching them as though they were a train passing through a station. Often patients report this simple experiment produces a marked decline in both the frequency of intrusions and the belief that they are a sign of impending insanity or loss of control.

Patients with social phobia often overestimate the significance of their anxiety symptoms for other people. A useful behavioural experiment to illustrate this point involves having either the patient or the therapist conduct a survey in which other people are asked for their views about the feared symptom. For example, in the case of fear of blushing, other people might be asked:

Why do you think people blush?
Do you notice other people blushing?
Do you remember it?
Do you think badly about people who blush?
If you do, what do you think about them?'

A further helpful experiment can involve intentionally displaying a feared symptom (e.g. handshaking or forgetting what one is talking about) and closely observing other people's responses. A particularly effective behaviour experiment for modifying social phobics distorted self-images involves the use of video feedback. Patients are asked to engage in a difficult social task while being videotaped. Afterwards they are asked to describe in detail how they think they appeared. They are then asked to view the video, watching themselves as though they are watching a stranger, ignoring memories of how they felt and simply focusing on how they would look to other people. In this way they often discover that they come across better than they would expect on the basis of their self-imagery. This experiment is often a powerful way of correcting distorted self-images.

Patients with panic disorder or hypochondriasis persistently think that normal bodily signs and/or symptoms are caused by a serious physical disorder. Numerous behavioural experiments can be used to demonstrate the correct, innocuous causes of their symptoms. For example, reading pairs of words which represent patient's illness interpretations (e.g. palpitations-dying, breathlessness-suffocate) has been shown to induce feared sensations.[5] Similarly, reproducing patients' fear-driven behaviours can produce the very symptoms the patients take as evidence for a serious physical illness. For example, patients who feel short of breath in a panic attack often respond by breathing quickly and deeply (hyperventilation), which paradoxically produces more breathlessness. Similarly, patients who are concerned about cancer may palpate body parts and then take the resulting soreness or discomfort as evidence of the presence of cancer.

(i) Therapy notes

Over a series of sessions therapist and patient will generate a substantial number of arguments against the patient's fearful beliefs. In order to maximize the impact of this accumulation, patients are asked to keep a running record of evidence against their beliefs in a notebook that can easily be consulted at times of doubt. Table 6.3.2.1.7 shows an illustrative example from a panic disorder patient's notebook. At the start of therapy, the patient had been concerned that there was something seriously wrong with his heart.

(j) Anger management

Although anxiety is the predominant problematic emotion in anxiety disorders, some patients also report significant problems with other emotions such as depression and anger. Techniques for dealing with depression can be found in Chapter 6.3.2.3. Some empirically validated techniques for dealing with anger are described here. Although presented in the context of anger accompanying anxiety disorders, these techniques are also relevant to anger in other disorders and to people without an Axis I disorder.

Table 6.3.2.1.7 A panic disorder patient's notebook: evidence for the two alternative explanations for chest pains

'There is something seriously wrong with my heart'	'My problem is my belief that there is something wrong with my heart'
1 *I hear my heart thumping sometimes, even in my ear.* But because of my fears I focus on my body and that makes me notice it. When I notice it I get anxious and that makes it louder because my heart beats are bigger	1 I think I am dying in a panic attack and that thought makes me anxious, producing many more sensations and setting up a vicious circle
2 *I have chest and rib tightness throughout the day.* But cardiac patients don't. They get chest pain (often crushing and more localized) during heart attacks. It is muscle tension due to work stress. It is mild after a good night's sleep and easier at weekends. It is worst after a stressful day at work	2 Distraction sometimes helps. That makes sense if the problem is my thoughts. It does not make sense if the problem is a heart attack. The same argument applies to leaving the situation. That would not stop a heart attack but it makes me feel more comfortable and undermines the negative thoughts
3 *I occasionally get tingling in my fingertips.* But this is a common symptom of anxiety. Also deep breathing—which I do when I think there is something wrong—causes tingling	3 I get symptoms most often at the end of the day, when I have come to expect them and have time to dwell on them
	4 I have proved to myself that there is nothing wrong with my heart with vigorous exercise. All that happens is that my heart beats faster and pumps harder, as it should do in order to supply my muscles with the energy they need

(k) Cognitive content and other assessment issues

Anger is triggered when other people are seen to have broken one's personal rules about what is right and fair.[1] Angry individuals invariably think that they have been badly treated and ascribe their perceived ill treatment to intention or unacceptable neglect on the part of others. A key first step in assessment is to help patients become aware of their automatic thoughts during periods of anger. It is also helpful to keep a record of the situations and behaviours of other people that routinely trigger anger. Review of such triggers often reveals a particular theme and an implicit rule that the patient thinks other people should abide by. A detailed description of how the person behaves when angry and what effect the behaviour has on others is also essential.

(l) Intervention

As patients' rules about the way that others should behave are often highly idiosyncratic, a useful tactic involves asking patients to consider whether the problem is assuming that others hold the same rule as them when they do not. This can help reduce the conviction that others' actions are actively malicious. Other useful questions include the following.

- Is there any other explanation for what happened?
- Did the other people know that their actions would harm me?
- Am I mind-reading?
- Am I over-applying the 'shoulds'?
- What are the advantages and disadvantages of responding with anger?
- Are there other ways I could behave which will be more likely to put things right/help me to get over it?

Although identifying and changing anger-related thoughts is a useful tactic, it is important to remember that anger is an action-orientated emotion. When angry, patients have a strong compulsion to hit out verbally or physically, and have great difficulty in thinking rationally. For patients with recurrent anger problems, it is often useful to teach them first to pause and relax or remove themselves from the anger-provoking situation before trying to challenge their thoughts and to delay-taking action (such as writing angry letters to others) until they have calmed down and had time to consider the appropriateness/usefulness of the action. To enhance further the generalizability of thought-challenging work, it is often useful to summarize the answers to typical anger-related thoughts on a flash card that patients can carry around and consult whenever they become angry.

Anger can sometimes be the result of chronic under-assertiveness, with patients' fears preventing them from making their point of view known until they feel overwhelmed and irritated by the demands placed on them. In such cases, discussion of the fears that prevent earlier and more appropriate assertion and role-playing in which the patients try out and evaluate ways of communicating their views to others in a prompt and constructive fashion can be helpful.

Indications and contraindications

Cognitive behaviour therapy is suitable for most patients with anxiety disorders and the low dropout rates reported in many controlled trials[4,32] suggest that it is well tolerated. In cases with additional severe comorbid problems (e.g. alcohol dependence, depression) it is sometimes necessary to bring these problems under control before starting cognitive behaviour therapy for anxiety. Concurrent use of prescription anxiolytic medication (benzodiazepines, tricyclics, selective serotonin reuptake inhibitors) is not a contraindication. At one time it was thought that anxiolytics may facilitate treatment by helping patients to confront their fears more quickly. However, there is little evidence that concurrent medication enhances initial response.[32] In addition, combining medication (alprazolam or imipramine) with cognitive behaviour therapy has been shown to produce poorer long-term outcome than cognitive behaviour therapy alone in panic disorder.[33] The latter result suggests that if a patient is not already taking anxiolytic medication, it is probably best to start treatment with cognitive behaviour therapy alone. Medication might then be added at a later stage, if response to cognitive behaviour therapy alone is poor.

Efficacy

Controlled trials involving comparisons with other psychological interventions and waiting-list control groups indicate that cognitive

behaviour therapy is an effective and specific treatment for panic disorder, social phobia, specific phobia, generalized anxiety disorder, hypochondriasis, obsessive–compulsive disorder, and post-traumatic stress disorder.[4,32] Results comparing immediate response to cognitive behaviour therapy alone and pharmacotherapy alone have been mixed, with superiority for cognitive behaviour therapy, equivalence for cognitive behaviour therapy, and superiority for pharmacotherapy all being reported. In contrast to the immediate response data, the follow-up analyses after medication discontinuation that are currently available favour cognitive behaviour therapy.[34] However, the database is modest and further research is required. For anger problems, controlled trials have shown that the cognitive behavioural procedures described here are effective.[35]

Training and supervision

Most controlled trials have used therapists who have received specialized training in cognitive behaviour therapy and there is some evidence that deviation from therapy protocols and/or poor implementation is associated with less good outcome.[36] For these reasons, clinicians are likely to benefit from specialized training and supervision. Where local training institutes exist, it is wise to take advantage of their expertise. Even when no local institute is available, expert cognitive behaviour therapists from established centres often travel internationally to deliver workshops and supervision. Several professional organizations run regular training workshops and can be contacted through the Internet. The organizations include the British Association of Behavioural and Cognitive Psychotherapies (http://www.babcp.org.uk), the Association of Behaviour and Cognitive Therapies (www.abct.org), the International Association of Cognitive Psychotherapy (http://www.cognitivetherapyassociation.org), the American Psychological Association (http://www.apa.org), and the American Psychiatric Association (www.psych.org). A comprehensive list of the competencies required for the main cognitive behaviour therapies for anxiety disorders can be found at: http://www.ucl.ac.uk/clinical-health-psychology/CORE/CBT_Framework.htm

Further information

A number of texts describe the theory and practice of cognitive behaviour therapy for specific anxiety disorders[37–39] and for anger problems[40] in considerable detail. Texts are frequently updated. Readers interested in the latest therapy guides are recommended to visit the following websites: www.oup.com, www.oup.com/us/ttw, www.guilford.com, www.wiley.com. Video illustrations of therapy sessions are also available for some anxiety disorders (see ABCT, American Psychological Association, and Guilford Press websites).

References

1. Beck, A.T. (1976). *Cognitive therapy and the emotional disorders*. International Universities Press, New York.
2. Rachman, S. (1996). The evolution of cognitive behaviour therapy. In *Science and practice of cognitive behaviour therapy* (eds. D.M. Clark and C.G. Fairburn), pp. 3–26. Oxford University Press, Oxford.
3. Marks, I.M. (1975). Behavioural treatment of phobic and obsessive-compulsive disorders. In *Progress in behavior modification* (eds. M. Hersen, R.M. Fisher, and P.M. Miller), **1**, pp. 66–158. Academic Press, New York.
4. Clark, D.M. (2004). Developing new treatments: on the interplay between theories, experimental science and clinical innovation. *Behaviour Research and Therapy*, **42**, 1089.
5. Clark, D.M. (1996). Panic disorder: from theory to therapy. In *Frontiers of cognitive therapy* (ed. P.M. Salkovskis), pp. 318–44. Guilford, New York.
6. APA. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, D.C.
7. Beck, A.T., Emery, G., and Greenberg, R.L. (1985). *Anxiety disorders and phobias: a cognitive perspective*. Basic Books, New York.
8. Borkovec, T.D., Ray, W.J., and Stober, J.W. (1998). A cognitive phenomenon intimately linked to affective, physiological, and interpersonal behavioral processes. *Cognitive Therapy and Research*, **22**, 561–76.
9. Wells, A. (2000). *Emotional disorders and metacognition: innovative cognitive therapy*. Wiley, Chichester.
10. Rachman, S.J. and De Silva, P. (1978). Abnormal and normal obsessions. *Behaviour Research and Therapy*, **16**, 233–48.
11. Salkovskis, P.M. (1985). Obsessional–compulsive problems: a cognitive-behavioural analysis. *Behaviour Research and Therapy*, **23**, 571–83.
12. Kessler, R.C., Sonnega, A., Bromet, E., *et al.* (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, **52**, 1048–60.
13. Ehlers, A. and Clark, D.M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, **38**, 319–45.
14. Brewin, C.R. and Holmes, E.A. (2003). Psychological theories of posttraumatic stress disorder. *Clinical Psychology Review*, **23**, 339–76.
15. Foa, E.B., Molnar, C., and Cashman, L. (1995). Change in rape narratives during exposure therapy for posttraumatic stress disorder. *Journal of Traumatic Stress*, **8**, 675–90.
16. Salkovskis, P.M. (1996). The cognitive approach to anxiety: threat beliefs, safety-seeking behaviour, and the special case of health anxiety and obsessions. In *Frontiers of cognitive therapy* (ed. P.M. Salkovskis), pp. 48–74. Guilford Press, New York.
17. Clark, D.M. and Wells, A. (1995). A cognitive model of social phobia. In *Social phobia: diagnosis, assessment and treatment* (eds. R. Heimberg, M. Liebowitz, D.A. Hope, and F.R. Schneier), pp. 69–93. Guilford Press, New York.
18. WHO. (1992). *International statistical classification of diseases and related health problems* (10th edn). WHO, Geneva
19. Veljaca, K.A. and Rapee, R.M. (1998). Detection of negative and positive audience behaviours by socially anxious subjects. *Behaviour Research and Therapy*, **36**, 311–21.
20. Hackmann, A., Surawy, C., and Clark, D.M. (1998). Seeing yourself through others' eyes: a study of spontaneously occurring images in social phobia. *Behavioural and Cognitive Psychotherapy*, **26**, 3–12.
21. Arntz, A., Rauner, M., and Van den Hout, M. (1995). If I feel anxious, there must be danger: ex-consequential reasoning in inferring danger in anxiety disorders. *Behaviour Research and Therapy*, **33**, 917–25.
22. Mansell, W. and Clark, D.M. (1999). How do I appear to others? Social anxiety and processing of the observable self. *Behaviour Research and Therapy*, **37**, 419–34.
23. Davey, G.C.L. and Matchett, G. (1994). Unconditioned stimulus rehearsal and the retention and enhancement of differential 'fear' conditioning: effects of trait and state anxiety. *Journal of Abnormal Psychology*, **103**, 708–18.
24. Beck, A.T. and Steer, R.A. (1993). *Beck anxiety inventory manual*. Psychological Corporation, San Antonio, TX.
25. Beck, A.T. and Steer, R.A. (1993). *Beck depression inventory*. The Psychological Corporation, San Antonio, TX.
26. Beunderman, R., Van Dis, H., Koster, R.W., *et al.* (1988). Differentiation in prodromal and acute symptoms of patients with cardiac and

non-cardiac chest pain. In *Advances in theory and practice in behaviour therapy* (eds. P.M.G. Emmelkamp, W.T.A.M. Everaerd, F. Kraaimaat, and M.J.M. van Son). Swets & Zeitlinger, Amsterdam.
27. Wells, A. and Hackmann, A. (1993). Imagery and core beliefs in health anxiety: content and origins. *Behavioural and Cognitive Psychotherapy*, **21**, 265–73.
28. Rachman, S.J., De Silva, P., and Roger, G. (1976). The spontaneous decay of compulsive urges. *Behaviour Research and Therapy*, **14**, 445–53.
29. Ehlers, A., Mayou, R.A., and Bryant, B. (1998). Psychological predictors of chronic posttraumatic stress disorder after motor vehicle accidents. *Journal of Abnormal Psychology*, **107**, 508–19.
30. NICE. (2005). *Post-traumatic stress disorder (PTSD): the management of PTSD in adults and children in primary and secondary care (Clinical Guideline 26)*. National Institute for Clinical Excellence, London, UK (www.nice.og).
31. Foa, E.B. and Riggs, D.S. (1993). Post-traumatic stress disorder in rape victims. In *Annual review of psychiatry* (eds. J.M. Oldham, M.B. Riba, and A. Tasman), pp. 273–303. American Psychiatric Association, Washington, DC.
32. Nathan, P.E. and Gorman, J.S. (eds.) (2002). *A guide to treatments that work*. Oxford University Press, New York.
33. Barlow, D.H., Gorman, J.M., Shear, M.K., *et al.* (2000). Cognitive-behavioral therapy, imipramine, or their combination for panic disorder. A randomized controlled trial. *Journal of the American Medical Association*, **283**, 2529–36.
34. Clark, D.M. and Wells, A. (1997). Cognitive therapy for anxiety disorders. In *Review of psychiatry* (eds. L.J. Dickstein, M.B. Riba, and J.M. Oldham), pp. 9–44. American Psychiatric Press, Washington, DC.
35. Del Vecchio, L. and O'Leary, K.D. (2004). Effectiveness of anger treatments for specific anger problems: a meta-analytic review. *Clinical Psychology Review*, **24**, 15–34.
36. Schulte, D., Kunzel, R., Pepping, G., *et al.* (1992). Tailor-made versus standardized therapy of phobic patients. *Advances in Behaviour Research and Therapy*, **14**, 67–92.
37. Hawton, K.E., Salkovskis, P.M., Kirk, J., *et al.* (1989). *Cognitive behaviour therapy for psychiatric problems*. Oxford University Press, Oxford.
38. Wells, A. (1997). *Cognitive therapy of anxiety disorders: a practice manual and conceptual guide*. Wiley, Chichester, UK.
39. Barlow, D.H. (ed.). (2007). *Clinical handbook of psychological disorders* (4th edn). Guilford Press, New York.
40. Howells, K. (1998). Cognitive behavioural interventions for anger, aggression and violence. In *Treating complex cases* (eds. N. Tarrier, A. Wells, and G. Haddock), pp. 295–318. Wiley, Chichester.
41. Spielberger, C.D., Gorsuch, R.L., and Lushene, R.E. (1970). *Manual for the state—Trait anxiety inventory*. Consulting Psychologists Press, Palo Alton, CA.
42. Clark, D.M., Salkovskis, P.M., Hackmann, A., *et al.* (1994). A comparison of cognitive therapy, applied relaxation and imipramine in the treatment of panic disorder. *British Journal of Psychiatry*, **164**, 759–69.
43. Marks, I. and Mathews, A.M. (1979). Brief standard self-rating for phobic patients. *Behaviour Research and Therapy*, **17**, 263–7.
44. Chambless, D.L., Caputo, G.C., Bright, P., *et al.* (1984). Assessment for fear of fear in agoraphobics: the Body Sensations Questionnaire and the Agoraphobia Cognitions Questionnaire. *Journal of Consulting and Clinical Psychology*, **52**, 1090–7.
45. Chambless, D.L., Caputo, G.C., Jasin, S.E., *et al.* (1985). The mobility inventory for agoraphobia. *Behaviour Research and Therapy*, **23**, 35–44.
46. Fresco, D.M., Coles, M.E., Heimberg, R.G., *et al.* (2001). The Liebowitz social anxiety scale: a comparison of the psychometric properties of self-report and clinical-administered formats. *Psychological Medicine*, **31**, 1025–35.
47. Mattick, R.P. and Clarke, J.C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, **36**, 455–70.
48. Turner, S.M., Beidel, D.C., Dancu, C.V., *et al.* (1989). An empirically derived inventory to measure social fears and anxiety: the social phobia and anxiety inventory. *Psychological Assessment*, **1**, 35–40.
49. Tallis, F., Davey, G.C.L., and Bond, A. (1994). The Worry Domains Questionnaire. In *Worrying: perspectives on theory, assessment and treatment* (eds. G. Davey and F. Tallis). Wiley, Chichester.
50. Molina, S. and Borkovec, T.D. (1994). The Penn State Worry questionnaire: psychometric properties and associated characteristics. In *Worrying: perspectives on theory, assessment and treatment* (eds. G. Davey and F. Tallis). Wiley, Chichester.
51. Wells, A. and Davies, M. (1994). The Thought Control Questionnaire: a measure of individual differences in the control of unwanted thoughts. *Behaviour Research and Therapy*, **32**, 871–8.
52. Salkovskis, P.M., Wroe, A.L., Gledhill, A., *et al.* (1999). Responsibility attitudes and interpretations are characteristic of obsessive compulsive disorder. *Behaviour Research and Therapy*, **38**, 347–72.
53. Van Oppen, P., Hoekstra, R.J., and Emmelkamp, P.M.G. (1995). The structure of obsessive-compulsive symptoms. *Behaviour Research and Therapy*, **33**, 15–23.
54. Goodman, W.K., Price, L.H., Rasmussen, S.A., *et al.* (1989). The Yale-Brown Obsessive-Compulsive Scale (YBOCS) Part 1: development, use and reliability. *Archives of General Psychiatry*, **46**, 1006–11.
55. Foa, E., Riggs, D.S., Dancu, C.V., *et al.* (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress*, **6**, 459–73.
56. Foa, E., Ehlers, A., Clark, D.M., *et al.* (1999). The post-traumatic cognitions inventory (PTCI): development and validation. *Psychological Assessment*, **11**, 303–14.
57. Horowitz, M.J., Wilner, N., and Alvarez, W. (1979). The Impact of Event Scale: a measure of subjective stress. *Psychosomatic Medicine*, **41**, 209–18.
58. Mechanic, M.B. and Resick, P.A. (1993). The personal beliefs and reactions scale: assessing rape-related cognitive schemata. *Paper presented at the 9th annual meeting of the International Society for Traumatic Stress Studies*, San Antonio, TX.
59. Clark, D.M., Ehlers, A., Hackmann, A., *et al.* (2006). Cognitive therapy and exposure plus applied relaxation in social phobia: a randomised controlled trial. *Journal of Consulting and Clinical Psychology*, **74**, 568–78.

6.3.2.2 Cognitive behaviour therapy for eating disorders

Zafra Cooper, Rebecca Murphy, and Christopher G. Fairburn

Introduction

The eating disorders provide one of the strongest indications for cognitive behaviour therapy. This bold claim arises from the demonstrated effectiveness of cognitive behaviour therapy in the treatment of bulimia nervosa and the widespread acceptance that cognitive behaviour therapy is the treatment of choice.[1] Cognitive behaviour therapy is also widely used to treat anorexia nervosa although this application has not been adequately evaluated. Recently its use has been extended to 'eating disorder not

otherwise specified' (eating disorder NOS),[2] a diagnosis that applies to over 50 per cent of cases,[3] and emerging evidence suggests that it is just as effective with these cases as it is with cases of bulimia nervosa.

In this chapter the cognitive behavioural approach to the understanding and treatment of eating disorders will be described. The data on the efficacy and effectiveness of the treatment are considered in the chapters on anorexia nervosa and bulimia nervosa (see Chapters 4.10.1 and 4.10.2 respectively), as is their general management.

The cognitive behavioural account of the maintenance of eating disorders

Although both the DSM and ICD schemes for classifying eating disorders encourage the view that anorexia nervosa and bulimia nervosa are distinct clinical states, consideration of their clinical features and course over time does not support this.[4] Patients with anorexia nervosa, bulimia nervosa, and eating disorder NOS have many features in common, most of which are not seen in other psychiatric disorders, and studies of their course indicate that most patients migrate between these diagnoses over time. This temporal movement, together with the fact that the disorders share the same distinctive psychopathology, has led to the suggestion that common 'transdiagnostic' mechanisms are involved in the persistence of eating disorder psychopathology.[5]

Anorexia nervosa, bulimia nervosa, and most cases of eating disorder NOS are united by a distinctive core psychopathology: patients over-evaluate the importance of their shape and weight and their ability to control them. According to the cognitive behavioural view it is this dysfunctional scheme of self-evaluation that is of central importance in maintaining these disorders. Whereas most people evaluate themselves on the basis of their perceived performance in a variety of domains of life, people with eating disorders judge themselves primarily in terms of their shape and weight and their ability to control them. Most of their other clinical features can be understood as stemming directly from this 'core psychopathology', including the extreme weight-control behaviour (i.e. the dieting, self-induced vomiting, laxative misuse, and over-exercising), the various forms of body checking and avoidance, and the preoccupation with thoughts about eating, shape, and weight. Fig. 6.3.2.2.1 provides a 'transdiagnostic' representation (or 'formulation') of the main processes involved in the maintenance of eating disorders.

The only feature that is not obviously a direct expression of the core psychopathology is binge eating, present in all cases of bulimia nervosa, many cases of eating disorder NOS and some cases of anorexia nervosa. The cognitive behavioural theory proposes that binge eating is largely a product of attempts to adhere to multiple extreme, and highly specific, dietary rules. These patients' tendency to react in a negative and extreme fashion to the (almost inevitable) breaking of these rules results in even minor dietary slips being interpreted as evidence of poor self-control. Patients respond to this perceived lack of self-control by temporarily abandoning their efforts to restrict their eating. This produces a highly distinctive pattern of eating in which attempts to restrict eating are repeatedly interrupted by episodes of binge eating. The binge eating maintains the core psychopathology by intensifying patients' concerns about their ability to control their eating, shape, and weight. It also encourages further dietary restraint, thereby increasing the risk of further binge eating.

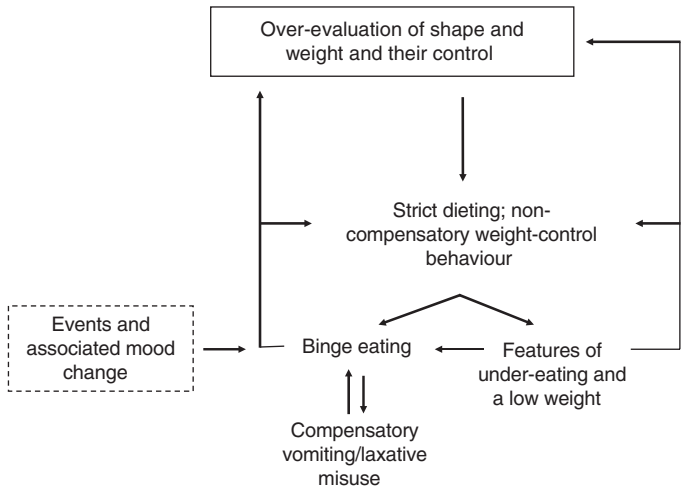


Fig. 6.3.2.2.1 The transdiagnostic 'template' formulation of the maintenance of eating disorders. (Reproduced from Fairburn, C.G., *Cognitive behavior therapy and eating disorders*, Copyright 2008, Guilford Press, NY).

Three further processes also maintain binge eating. First, difficulties in the patient's life and associated mood changes increase the likelihood that they will break their dietary rules. Second, binge eating temporarily ameliorates such mood states and distracts patients from thinking about their difficulties. Third, if the binge eating is followed by compensatory vomiting or laxative misuse, it is also maintained because patients' mistaken belief in the effectiveness of such 'purging' undermines a major deterrent to their binge eating. They do not realize that purging has little effect on energy absorption.[6]

Patients with anorexia nervosa share the distinctive core psychopathology of those with bulimia nervosa and eating disorder NOS. The major difference between patients with anorexia nervosa and those with other eating disorders lies in the fact that in anorexia nervosa under-eating predominates and therefore patients become extremely underweight. This has certain physiological and psychological consequences (see Chapter 4.10.1) that contribute to the persistence of the eating disorder.[7] For example, delayed gastric emptying results in a sense of fullness even after eating modest amounts of food and secondary social withdrawal magnifies patients' isolation from the influence of others.

The composite 'transdiagnostic' formulation shown in Fig. 6.3.2.2.1 represents the core processes that maintain any eating disorder. The specific maintaining processes operating in any patient depend upon the nature of the eating disorder psychopathology present. In some cases only certain of the processes are active (for example, in most cases of binge-eating disorder), but in others (for example, cases of the binge eating/purging subtype of anorexia nervosa) most are operating. The formulation highlights the maintaining processes that need to be addressed in treatment, thereby allowing the clinician to design a bespoke treatment to fit the individual patient's psychopathology.

Evidence for the cognitive behavioural account

There is a sizeable body of research that supports the cognitive view of the maintenance of eating disorders.[8] This includes descriptive and experimental studies of the clinical characteristics of these

patients and the research on dietary restraint and 'counter-regulation' (a possible analogue for binge eating).[9] There is also strong indirect support from a large body of research indicating that cognitive behaviour therapy based on this model has a major and lasting impact on bulimia nervosa (see Chapter 4.10.2) and emerging evidence that this is also true of eating disorder NOS. Further support comes from the finding that 'dismantling' the cognitive behavioural treatment for bulimia nervosa by removing those procedures designed to produce cognitive change attenuates its effects and results in patients being markedly prone to relapse.[10] Direct support comes from studies that have shown that dietary restraint appears to mediate this treatment's effect on binge eating[11] and that continuing over-evaluation of weight and shape in those who have recovered in behavioural terms is predictive of subsequent relapse.[12,13]

Transdiagnostic cognitive behavioural treatment of eating disorders

There follows a brief description of cognitive behaviour therapy for eating disorders. Full details are provided elsewhere.[14] The treatment is outpatient-based and, as applied in research settings, involves 20 individual treatment sessions over 20 weeks for the 80 per cent or more of patients who are not significantly underweight (body mass index over 17.5). The remaining patients with a body mass index of 17.5 or below receive 40 sessions over 40 weeks. The indications for the treatment are the presence of an eating disorder of clinical severity. It is not appropriate for those whose psychiatric state, general physical health, or degree of weight loss is such that they cannot safely be treated on an outpatient basis.

The treatment has four stages.

(a) Stage one

The aims of the first stage are as follows: to educate patients about treatment and the disorder; to engage the patient in treatment and change, and to introduce and establish a pattern of regular eating and weekly weighing. This stage comprises approximately eight sessions which are held twice weekly over 4 weeks.

(i) Jointly creating the formulation

This is usually done in the first treatment session. The therapist draws out the relevant sections of Fig. 6.3.2.2.1 incorporating the patient's own experiences and terms. This helps patients to realize both that their behaviour is comprehensible and that it is maintained by a variety of self-perpetuating mechanisms which are open to change. The formulation provides a guide to what needs to be targeted in treatment if patients are to achieve a full and lasting recovery.

(ii) Establishing real-time self-monitoring

This is the ongoing 'in-the-moment' recording of eating and other relevant behaviour, thoughts, feelings, and events (see Chapter 4.10.2 for an example monitoring record). Self-monitoring is initiated in the first session and continues throughout treatment. It serves two purposes: it assists in the identification of the patient's problems and progress and, more importantly, it facilitates change by helping patients address problems as they occur.

(iii) Establishing 'weekly weighing'

The patient and therapist check the patient's weight once a week and plot it on an individualized weight graph. Patients are strongly encouraged not to weigh themselves at other times. Weekly in-session weighing has several purposes: first, it provides patients with accurate data about their weight at a time when their eating habits are changing; second, it provides an opportunity for the therapist to help patients interpret the numbers on the scale, which otherwise they are prone to misinterpret and, third, it addresses the important maintaining processes of excessive body weight checking or its avoidance.

(iv) Providing education

From the second session onwards, an important element of treatment is education about weight and eating since many patients have misconceptions that maintain their eating disorder. The following topics need to be covered:

- Body weight and its regulation: the body mass index and its interpretation; natural weight fluctuations; and the effects of treatment on weight.
- Physical complications of binge eating, self-induced vomiting, the misuse of laxatives and diuretics, and the effect of the eating disorder on hunger and fullness.
- Ineffectiveness of vomiting, laxatives, and diuretics as a means of weight control.
- Adverse effects of dieting: the types of dieting that promote binge eating; dietary rules versus dietary guidelines.

To provide reliable information on these topics, patients are asked to read relevant sections from one of the authoritative books on eating disorders[6,15,16] and their reading is discussed in subsequent treatment sessions.

(v) Establishing 'regular eating'

The establishment of a pattern of regular eating is fundamental to successful treatment whatever the form of the eating disorder. It addresses an important type of dieting ('delayed eating'); it displaces episodes of binge eating and, for underweight patients, it introduces regular meals and snacks that can be subsequently increased in size. Early in treatment (usually by the third session) patients are asked to eat three planned meals each day, plus two (or if underweight three) planned snacks and they are asked not to eat between them. Patients may choose what they eat at these times with the only conditions being that the meals and snacks are not followed by any compensatory behaviour and that there should rarely be more than a 4-hour interval between these occasions of eating. The new eating pattern should take precedence over other activities but should not be so inflexible as to preclude the possibility of adjusting timings to suit the patients' commitments each day.

Patients should be helped to adhere to their regular eating plan and to resist eating between the planned meals and snacks. Two rather different strategies may be used to achieve this: the first involves helping patients to identify activities that are incompatible with eating or make it less likely, and the second is to help patients to recognize that the urge to eat is a temporary phenomenon. Through using these strategies patients learn to distance themselves from the urge to eat which they find gradually fades with time.

(vi) Involving significant others

The treatment is primarily an individual treatment for adults and hence it does not actively involve others. Despite this, it is our

practice to see 'significant others' with the patient if this is likely to facilitate treatment and the patient is willing for this to happen. There are two specific indications for involving others: if others could help the patient in making changes or if others are making it difficult for the patient to change by, for example, commenting adversely on eating or appearance.

(b) Stage two

Stage two is a transitional stage which generally comprises two appointments, a week apart. Whilst continuing with the procedures introduced in stage one the therapist and patient conduct a joint review of progress to date, identify problems still to be addressed, revise the formulation if necessary, and design stage three.

(c) Stage three

The aim of this stage is to address the key mechanisms that are maintaining the patient's eating disorder. The order in which these mechanisms are addressed depends upon their relative importance in maintaining the particular patient's psychopathology. There are generally 8-weekly appointments.

(i) Addressing the over-evaluation of shape and weight

The first step involves explaining the concept of self-evaluation and helping patients identify the life domains which contribute to their judgement of themselves. The relative importance of these domains can be visually represented on a pie chart, which for most patients is dominated by a large slice representing shape and weight and controlling eating.

The patient and therapist then identify the problems inherent in this scheme for self-evaluation. Briefly there are three related problems: first, the over-evaluation of shape and weight tends to marginalize other domains and thus self-evaluation is overly dependent on performance in one area of life; second, the area of controlling shape and weight is one in which success is elusive, thus undermining self-esteem; and third, the over-evaluation leads to behaviour which is unhelpful and which itself maintains the disorder.

The final step in educating about self-evaluation involves identifying its three main expressions which occur to varying degrees in different patients. These are body checking, body avoidance, and feeling fat. The therapist explains how these behaviours and experiences serve to maintain and magnify the patient's concerns about shape and weight and it is agreed therefore that they need to be addressed in treatment.

(ii) Addressing body checking and avoidance

Patients are often not aware that they are engaging in body checking and that it is maintaining their body dissatisfaction. The first step in addressing body checking involves obtaining a detailed account of the behaviour by asking patients to record it. An example monitoring record is shown in Fig. 6.3.2.2.2. Patients are then helped to realize that body checking is not a helpful way of assessing their

Day ……Friday…….. Date …..6th Sept ……..

Time	Food and drink consumed	Place	*	V/L	Checking (what done, time taken)	Place	Context and comments
7.45	1 piece of toast with butter and marmite. 1 cup of tea	Kitchen					Not hungry but know I should have breakfast
8.15					Looking at stomach and thighs in bedroom mirror while getting dressed (2 mins)	Bedroom	Depressing. Can't see any muscles – only fat.
10.30	1 apple & diet coke	Office					
11.15					Scrutinisng stomach in mirror standing sideways (1 min)	Office toilet	Ok
			*				How can it be so fat?? Have hardly eaten anything today!
12.30	Cheese and tomato sandwich & banana and kit kat	Canteen					Chocolate was too much. Feel too full.
1.30					Feeling and pinching stomach while sitting at desk (10mins)	Office	myself sick.
							Fat disgusting flesh. Feel massive.
3.10	Cup of coffee and 1 yoghurt	Office					Had planned to have other 1/ 2 of chocolate bar. Can't do it, am already too fat!
3.30	Water				Assessing shape of woman on street (10secs)	Office – through window	Feel unhappy. Wish I was as thin as her. Can't accurately compare myself to others – never really sure what my shape is. Frustrating.
6.45	Salad with tuna 1 glass of red wine	Living room					
7.30					Looking at reflection in window while doing dishes (5mins)	Kitchen	Still want more food but won't let myself since might lose control.
						Bathroom	
8.10			*	V	Touching and pinching stomach and thighs while having a bath (15mins)		I'm so huge! Wish I was tall and elegant. How depressing.
9.30	1 kit kat and half a crunchie 1 glass of red wine	Living room					Feel disgusted. Am I ever going to get rid of this fat?
							Feel fat! Too much chocolate today. Have to get rid of the food. Go to bed to stop thinking about it.

Fig. 6.3.2.2.2 An adapted monitoring record illustrating body checking.

shape or weight as it provides unreliable and biased information. Certain forms of body checking are best stopped altogether. In the case of more normative checking such as mirror use, education should stress that, as with other forms of body checking, what one finds depends to an important extent upon how one looks (e.g. scrutiny of perceived flaws tends to magnify them). For patients who avoid seeing their bodies, the therapist needs to explain that this too maintains dissatisfaction. Patients need to be encouraged to get used to the sight and feel of their body. Participation in activities that involve a degree of body exposure can be helpful, for example, swimming.

(iii) Addressing 'feeling fat'

'Feeling fat' is an experience reported by many women but the intensity and frequency of this feeling appears to be far greater among people with eating disorders. Feeling fat is a target for treatment since it tends to be equated with being fat (irrespective of actual shape and weight) and hence maintains body dissatisfaction. Although this topic has received little research attention, clinical observation suggests that in many patients feeling fat is a result of mislabelling certain emotions and bodily experiences. It may be addressed by helping patients appreciate that feeling fat tends to be triggered by the occurrence of certain negative mood states (e.g. feeling bored or depressed) or by physical sensations that heighten body awareness (e.g. feeling full, bloated, or sweaty). Patients can then be encouraged to question the feeling when it occurs and correctly label and address the underlying triggering state using a problem-solving approach.

(iv) Developing marginalized domains for self-evaluation

Tackling the expressions of the over-evaluation of shape and weight will gradually reduce it. At the same time, it is also important to encourage the patient to increase the number and significance of other domains for self-evaluation. Although this is an indirect means of diminishing the over-evaluation of shape and weight, it is nevertheless a powerful one.

(v) Exploring the origins of the over-evaluation

Towards the end of stage three it is often helpful to explore the origins of the patient's sensitivity to shape, weight, and eating. An historical review can help to make sense of how the problem developed and evolved, highlight how it might have served a useful function in its early stages and help patients distance themselves from the past. If a specific event appears to have played a critical role in the development of the eating problem, the patient should be helped to reappraise this from the vantage point of the present.

(vi) Addressing dietary restraint

A major goal of treatment is to reduce, if not eliminate altogether, strict dieting. This dieting has two aspects: an attempt to limit eating termed 'dietary restraint' and actual under-eating in physiological terms termed 'dietary restriction'. 'Regular eating' will already have addressed one form of dietary restraint (delayed eating). Patients need to recognize that their multiple extreme and rigid dietary rules lead to preoccupation with food and eating, encourage binge eating and impose practical and social restrictions. It should therefore be agreed that dietary restraint needs to be addressed. To do this, the patient's various dietary rules should be identified together with the beliefs which underlie them. The patient should be helped to break the rules in order to test the beliefs in question and to learn that the feared consequences that maintain the dietary rule (typically sudden weight gain or binge eating) are not an inevitable result of breaking it. With patients who binge eat it is important to pay particular attention to food avoidance and to help them systematically reintroduce such foods into their diet.

(vii) Addressing event-triggered changes in eating

Among patients with eating disorders, eating habits may change in response to outside events. The change may involve eating less, stop eating altogether, overeating or binge eating. If these changes persist into stage three, they should be addressed by helping patients to tackle the triggering events using a problem-solving approach and by helping patients to accept the occurrence of intense-mood states and identify ways (that are not harmful) of modulating their moods.

(d) Stage four

The aims in stage four are to ensure that the changes made in treatment are maintained over the following months and that the risk of relapse is minimized in the long term. There are three appointments, each 2 weeks apart. During this stage, as part of their preparation for the future, patients discontinue self-monitoring and transfer from in-session weighing to weighing themselves at home.

To maximize the chances that progress is maintained the therapist and patient jointly devise a specific plan for the patient to follow over the following few months until a post-treatment review appointment. Typically this includes further work on body checking, food avoidance, and perhaps further practice at problem-solving. In addition, the therapist encourages patients to continue their efforts to develop new interests and activities.

There are two elements to 'relapse prevention'. First, patients must have realistic expectations regarding the future. A common problem is that many hope never to experience any eating difficulties again. It needs to be explained that this makes them vulnerable to relapse since it encourages a negative reaction to even minor setbacks. Patients should be told to expect lapses with the eating problem continuing to be their Achilles' heel. The goal is for patients to identify setbacks as early as possible, view them as a 'lapse' rather than a 'relapse', and use a well-developed plan to deal with them. Thus, the second element of relapse prevention is the construction of such a plan. The therapist and patient should review the components of treatment with the aim of identifying the principles and procedures that were most relevant and helpful and devise a plan for the future incorporating this information.

Underweight patients

When treating patients who are underweight (most are cases of anorexia nervosa but some are cases of eating disorder NOS) three main modifications to the treatment are required: the motivation of these patients needs to be enhanced, their state of starvation needs to be corrected, and significant others are more likely to be involved. As a result treatment needs to be considerably longer.

(a) Enhancing motivation to change

The poor motivation of these patients needs to be addressed from the outset of treatment. There are various ways of enhancing motivation[17] including focusing on establishing a sound therapeutic relationship, ensuring that the patient feels understood, making it clear that one is working on behalf of the patient and not their

relatives or concerned others, accepting the patient's beliefs and values as genuine and comprehensible, and adopting an experimental approach in which the therapist and patient together explore the advantages and disadvantages of making changes. This includes educating the patient about the physiological and psychological effects of starvation; for example, impaired concentration, preoccupation with food and eating, sleep disturbance, sensitivity to cold, ritualistic eating, social withdrawal, and enhanced fullness secondary to delayed gastric emptying.[7] It is best to focus particularly on those features that the patient views as a problem and explain how they tend to perpetuate the eating disorder. In addition, an exploration of the broader impact of the eating problem on the patient's life is important. When exploring the advantages and disadvantages of change, it is important to draw a distinction between the short-term and long-term consequences of change since patients tend to focus on the immediate present rather than the future.

(b) Restoring a healthy weight

Unless the weight loss is rapid or extreme, or the patient's health is endangered by physical complications, weight restoration can usually be accomplished on an outpatient basis. Before focusing on weight gain, however, it is best to devote several sessions to establishing a collaborative working relationship and to developing a joint formulation and treatment plan. Thereafter, weight gain and the subsequent maintenance of a healthy weight must be an integral part of treatment. A target weight range should be identified in excess of a body mass index of 19.0.

The weight gain should be gradual and steady (at an average rate of about 0.5 kg/week). This requires an energy surplus to be established. This can be achieved by providing patients with energy-rich drinks to supplement their food intake (which should be increased such that it is sufficient to maintain their current weight). The energy-rich drinks may be viewed as weight restorative 'medicine' designed to produce the energy surplus. The drinks should be phased out once the target weight range has been reached. Whilst regaining weight patients should be helped to address their shape and weight concerns and dieting in much the same way as described above. Once a satisfactory weight has been reached patients need time to learn how to maintain their weight in a normal manner.

(c) Involving the significant others

Weight regain is a protracted process that takes considerable time and effort on the part of the patient. If the patient lives with others it can be helpful to involve them if doing so is consistent with the nature of their relationship. Significant others can help the patient choose what to eat and provide guidance concerning portion sizes. This is likely to be especially important with younger patients (many underweight patients are adolescents) who are still living at home with their parents.

Acknowledgements

We are grateful to the Wellcome Trust for its support. CGF holds a Principal Research Fellowship (046386). ZC and RM are supported by a programme grant (046386).

Further information

Fairburn, C.G. (2008). *Cognitive behavior therapy and eating disorders*. Guilford Press, New York.

The International Journal of Eating Disorders.

References

1. National Institute for Clinical Excellence. (2004). *Eating disorders—core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders*. NICE Clinical Guideline No. 9. NICE, London. www.nice.org.uk
2. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*. American Psychiatric Association, Washington, DC.
3. Fairburn, C.G. and Bohn, K. (2005). Eating disorder NOS (EDNOS): an example of the troublesome 'not otherwise specified' (NOS) category in DSM-IV. *Behaviour Research and Therapy*, **43**, 691–701.
4. Fairburn, C.G. and Harrison, P.J. (2003). Eating disorders. *Lancet*, **361**, 407–16.
5. Fairburn, C.G., Cooper, Z., and Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: a 'transdiagnostic' theory and treatment. *Behaviour Research and Therapy*, **41**, 509–28.
6. Fairburn, C.G. (1995). *Overcoming binge eating*, pp. 48–54. Guilford Press, New York.
7. Garner, D.M. (1997). Psychoeducational principles in treatment. In *Handbook of treatment for eating disorders* (eds. D.M. Garner and P.E. Garfinkel), pp. 145–77. Guilford Press, New York.
8. Vitousek, K.M. (1996). The current status of cognitive–behavioral models of anorexia nervosa and bulimia nervosa. In *Frontiers of cognitive therapy* (ed. P. Salkovskis), pp. 383–418. Guilford Press, New York.
9. Polivy, J. and Herman, C.P. (1993). Etiology of binge eating: psychological mechanisms. In *Binge eating: nature, assessment and treatment* (eds. C.G. Fairburn and G.T. Wilson), pp. 173–205. Guilford Press, New York.
10. Fairburn, C.G., Jones, R., Peveler, R.C., *et al.* (1993). Psychotherapy and bulimia nervosa: the longer–term effects of interpersonal psychotherapy, behaviour therapy and cognitive behaviour therapy. *Archives of General Psychiatry*, **50**, 419–28.
11. Wilson, G.T., Fairburn, C.G., Agras, W.S., *et al.* (2002). Cognitive-behavioral therapy for bulimia nervosa: time course and mechanisms of change. *Journal of Consulting and Clinical Psychology*, **70**, 267–74.
12. Fairburn, C.G., Peveler, R.C., Jones, R., *et al.* (1993). Predictors of 12–month outcome in bulimia nervosa and the influence of attitudes to shape and weight. *Journal of Consulting and Clinical Psychology*, **61**, 696–8.
13. Fairburn, C.G., Stice, E., Cooper, Z., *et al.* (2003). Understanding persistence in bulimia nervosa: a 5-year naturalistic study. *Journal of Consulting and Clinical Psychology*, **71**, 103–9.
14. Fairburn, C.G. (2008). *Cognitive behavior therapy and eating disorders*. Guilford Press, New York.
15. Cooper, P.J. (1995). *Bulimia nervosa and binge eating: a guide to recovery*. Robinson, London.
16. Schmidt, U. and Treasure, J. (1993). *Getting better bit(e) by bit(e)*. Erlbaum, Hove.
17. Vitousek, K., Watson, S., and Wilson, G.T. (1998). Enhancing motivation for change in treatment–resistant eating disorders. *Clinical Psychology Review*, **18**, 391–420.

6.3.2.3 Cognitive behaviour therapy for depressive disorders

Melanie J. V. Fennell

Introduction

This chapter describes A.T. Beck's cognitive behaviour therapy (CBT) for depression.[1] Beck's is probably the most fully developed, comprehensively evaluated, and widely disseminated cognitive behavioural approach to depression. Additionally, CBT is an effective treatment for a range of acute psychiatric disorders, shows promise for severe mental illness and personality disorder, and is thus helpful not only with primary depression, but also with a range of comorbid conditions.

Central characteristics of CBT

The general principles and nature of CBT are described elsewhere. Two specific points relate to depression:

(a) Demands of CBT

Given the nature of depression, CBT challenges both therapist and patient. It requires **active engagement** (e.g. willingness to carry out self-help assignments), yet depressed patients often lack motivation and energy. It is based on a **friendly collaboration**, yet depressed people often find it hard to talk and clinicians may find their negativity aversive. It is an **educational approach**, using written materials and record sheets, yet depressed patients often have concentration and memory difficulties. Its stance is **optimistic**, yet depressed patients are often afraid (or convinced) that change is impossible. Therapists should be alert to these difficulties, understand them as aspects of depression rather than blaming the patient ('She must want to be depressed') or themselves ('I'm no good at this'), and maintain a persistent, problem-solving stance.

(b) Advantages of CBT

Nonetheless, CBT has real advantages for depressed patients. Its **structure** discourages rumination, and helps patients to focus systematically on their difficulties. Its emphasis on a warm **therapeutic relationship** encourages empathy, while its **goal orientation** implies that change *is* possible. The **coherent model** of human functioning, on which it is based, allows it to address many issues, including depression itself, comorbid conditions, problems in living, long-standing difficulties (such as low self-esteem), patients' responses to therapy and therapist, and therapists' responses to patients. Its **emphasis on collaboration and on transfer of knowledge and skill** empowers patients to become their own therapists and to take control of their lives.

Background

Beck's interest in the role of cognition grew out of his practice as a psychoanalytical therapist. Dissatisfied with analytical understandings of depression, he became curious about the role of the negative thinking he observed in depressed patients.[2–4] Beck's clinical observation and research consistently showed the thinking of depressed people to be dominated by self-derogation, negative expectations, overwhelming problems and responsibilities, deprivation and loss, and escapist and suicidal wishes, themes fuelled by systematic biases in information processing. He suggested that patients could recover from depression by learning to re-evaluate everyday cognitions, and to understand the long-standing idiosyncratic schemas under lying them.

This early scientist–practitioner stance has remained central CBT, stimulating an integrated flow of experimental investigation, practice development, and research into treatment efficacy, which has continued to the present day. The first successful outcome trial of CBT for depression appeared in 1977.[5] A detailed treatment manual emerged shortly afterwards.[1] Thirty years of randomized controlled trials now support the treatment's effectiveness.[6] Like other short-term focused psychological treatments, it has consistently proved as effective as antidepressant medication post-treatment. It reduces the likelihood of relapse by about 50 per cent, and this effect endures,[7] and can be enhanced by booster sessions in the months following treatment.[8] Thus CBT emerges as surprisingly cost-effective.[9] With adequate training and supervision equivalent results can be achieved even with severe depression and high comorbidity.[10,11]

Technique

Cognitive case conceptualization

(a) Enduring cognitive vulnerability to depression

The **cognitive model of depression** proposes that enduring cognitive structures and processes shape how everyday experience is interpreted, and are in turn reinforced by these interpretations. This model (Fig. 6.3.2.3.1) forms the basis for an **individualized conceptualization**, developed and shared with the patient, which informs and guides therapy. It suggests that **experience** (loss, events with lasting implications for self-worth)[12,13] leads people to reach **fundamental conclusions** about themselves, others, and the world ('basic' or 'core' beliefs, or schemas). They devise **guidelines for living** ('conditional assumptions'), which allow them to operate in the world, assuming the truth of those conclusions. Using schemas and rules to organize experience and guide behaviour is a normal part of human functioning. However, where schemas are globally negative (e.g. 'I am inferior') and assumptions extreme and resistant to change (e.g. stringent perfectionism), they become counterproductive. Evidence for **cognitive vulnerability** to depression prior to first onset is now emerging,[14] as is evidence that recurrent episodes leave an tendency to re-experience depressogenic processing patterns in the presence of mild, normal depressed mood ('cognitive reactivity').[15,16]

(i) Relationship between thinking and other aspects of depression

Dysfunctional beliefs and assumptions are **activated by events** that match the person's particular sensitivities. So a person with negative beliefs about the self whose psychological well-being depends on love and approval might become depressed after experiencing rejection. Activation of the system results in an upsurge of '**negative automatic thoughts**'—'negative' in that they are associated with painful emotions, and 'automatic' in that they pop into the person's mind rather than being a product of reasoned reflection. Such thoughts reflect **processing biases** such as overgeneralization. Depression is characterized by biased negative thoughts about the self (e.g. 'I'm useless'), the world (e.g. 'My situation is intolerable'), and the future (e.g. 'Nothing will ever change'). The latter (hopelessness) is central to suicidality.

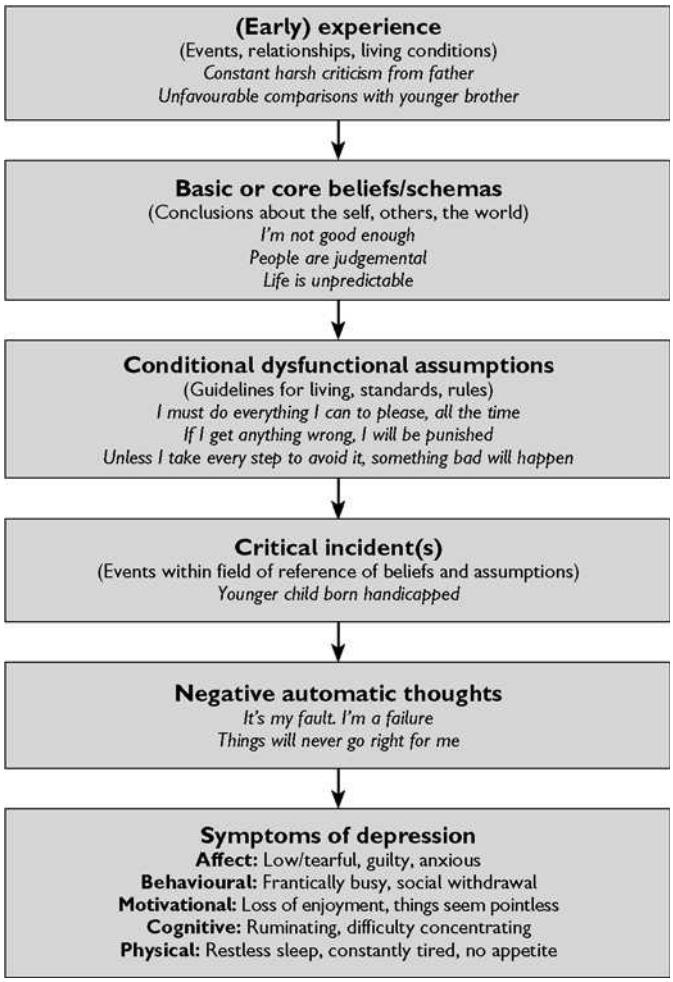


Fig. 6.3.2.3.1 Cognitive model of depression.

The more depressed people become, the more negative thoughts they think, and the more they believe them. The more negative thoughts they think and the more they believe them, the more depressed they become. Thus depression is maintained by **vicious circles** in which negative thinking and other symptoms reinforce one another. Experimental and clinical research reflects this **reciprocal relationship between affect and cognition**: modifying depressive thinking modifies depressed mood, while modifying depressed mood modifies depressive thinking.

(ii) Objectives of CBT for depression

The first goal is to **break the vicious circle** described above by teaching patients to work effectively with negative automatic thoughts. Attention then turns to cognitive predisposing factors (beliefs, assumptions) in order to **reduce vulnerability** and decrease the likelihood of future episodes. The aim is not to teach realistic thinking *per se*, but rather to help patients to resolve their problems by teaching them cognitive behavioural skills.

(iii) Overview of CBT for depression

Treatment usually proceeds through the following **stages**:

- diagnosis, assessment, problem identification
- cognitive interventions designed to reduce the frequency of negative thoughts
- behavioural assignments intended to tackle behavioural and motivational deficits
- monitoring, questioning, and testing negative automatic thoughts (the main body of therapy)
- relapse prevention.

Patients usually move from one stage to the next as each is mastered. The **starting point varies**: severely depressed patients often begin with simple behavioural interventions, whereas relatively mild depressions may immediately be amenable to cognitive work. At each stage, **cognitive and behavioural interventions are closely integrated**.[17] Thus behavioural interventions such as activity scheduling present opportunities to identify, question, and test negative thoughts, while work on thoughts and assumptions is closely linked to changes in behaviour.

Traditionally, **up to 20 sessions** of therapy are offered, twice weekly for 4 weeks and once weekly thereafter. Most patients respond within about 15 sessions. Some do well with 4–6 sessions, but severe chronic depressions may require more than 20 sessions, as well as shorter, more frequent sessions early on. Post-treatment boosters help to increase confidence and consolidate and extend skill. Sessions start with agenda-setting (prioritizing what to work on), homework review, and feedback on the previous session. After the day's main topics have been discussed, more homework practice is agreed to ensure generalization and reinforce new learning. Key points are summarized, and the therapist asks for reactions to the session, including anything that has been uncomfortable or unclear.

Indications and contraindications

CBT was developed as a treatment for moderate-to-severe unipolar depression,[1] and has been consistently effective with this population. However, good average results in outcome trials conceal wide variations in responsiveness.[18] Consistent predictors of treatment response remain elusive.

Patients presenting with endogenous symptoms are as likely to respond well as non-endogenous patients. However, results for severe depression[19] and bipolar disorder[20] remain somewhat conflicting.

Some factors may facilitate a positive response; well-developed pre-existing cognitive and behavioural coping skills, acceptance of the cognitive model, willingness to engage in self-help assignments, an early focus on teaching specific cognitive-therapy skills, ease of access to thoughts and feelings, problem specificity, and ability to form a collaborative alliance. Such criteria can be explored at assessment, especially if it is divided into two sessions with a simple intervening homework assignment.

Patients who are unsuitable for short-term CBT[21] may nonetheless respond to a modified approach with:

- greater emphasis on socialization and on cultivating a solid working relationship
- more extended behavioural work
- more work on enduring depressogenic schemas in the later stages of treatment
- more emphasis on integration with antidepressant medication
- more careful attention to environmental factors (including not only life stresses and family relationships, but also the ward and health-care team)

Finally, given that CBT and antidepressant medication generally produce similar results, the patients' wish for psychological treatment should also be taken into account.

Selection procedure

Diagnosis: recognizing and labelling depression

CBT was designed for moderate-to-severe major unipolar depression. The diagnostic criteria for major depressive episode,[22] remarkably consistent for over 30 years, describe a symptom pattern (including cognitive features) that has been recognized throughout history, and appears basically consistent across age, gender, race, and culture. However, the relative emphasis on different symptoms and the manner in which distress is expressed varies, and assessment procedures should be adapted to explore cultural context. Additionally, sociocultural factors necessarily influence belief systems, and therapists should be sensitive to such differences, rather than assuming that their own assumptions are shared by their patients.

Severity

Severity should be taken into account when deciding whether concurrent (or alternative) physical treatments or hospitalization are necessary, and in determining where to begin within CBT. Severity can be assessed through clinical interview (e.g. intensity, pervasiveness, and reactivity of depressed mood; extent of behavioural and interpersonal deficits). The **Beck Depression Inventory** (BDI-II)[23] a well-established self-report measure of depression, provides a rapid overview of symptoms. Weekly completion allows clinicians to observe overall progress, as well as tracking scores on particular items, e.g. hopelessness. Hopelessness and suicidality should be routinely assessed[24] and any sign of suicide risk investigated in-depth.[25]

Clinical interview

Interview assessment for CBT includes the following:

(a) Symptoms and associated cognitions

Negative automatic thoughts both trigger and enhance symptoms of depression (Table 6.3.2.3.1). Identifying meanings attached to symptoms prepares the ground for more helpful perspectives (e.g. 'These are symptoms of depression, not a reflection of my worth as a person').

(b) Impact on functioning

It is important to establish how depression affects relationships, work performance, and leisure time. It may be necessary to take practical steps to improve the patient's situation (e.g. gradual reintroduction to work).

(c) Coping strategies

The more depressed the patient, the more likely that she/he has adopted coping strategies which help in the short-term, but are in the longer-term self-defeating (e.g. alcohol or drugs, social withdrawal, bed). Therapist and patient can discuss the pros and cons of these, and how cognitive behavioural strategies might be more beneficial. The aim is for patients to reach the point of trying more adaptive coping strategies for themselves.

Table 6.3.2.3.1 Negative automatic thoughts and symptoms of depression

Symptoms	Negative automatic thoughts
Behavioural	
Lowered activity level	I can't do it. It won't work.
Procrastination	I'll never get it done. It's too much for me.
Motivational	
Loss of energy	It's too much effort. I'll wait till I fell better.
Loss of pleasure, interest	I won't enjoy it. What's the point? I can't be bothered.
Affective	
Sadness	I've lost everything important to me.
Guilt	I'm letting everybody down.
Anger	Why can't people just leave me alone?
Shame	What must everyone think?
Anxiety	I'm not going to be able to cope.
Cognitive	
Indecisiveness	Whatever I do will go wrong.
Poor concentration	I must be going senile.
Physical	
Loss of sleep	If this goes on, I won't be able to function.
Loss of appetite	I'm going to make myself ill.
Loss of sexual appetite	Our marriage is at an end.
Other	
Suicide	This is unbearable. There is no other way out.
Problems in living	There's nothing to be done.

(d) Onset of current episode

Information about the onset of episodes may provide valuable clues about beliefs and assumptions. For example, a young woman became depressed when her husband took a job abroad, and was away for long periods. She believed that he would not have done so if he truly loved her. In fact, he had taken the job because it paid exceptionally well and the savings they could make would allow them to start the family they had been planning. The therapist noted the patient's interpretation of her husband's behaviour, and later in therapy used this clue as a starting point to identify long-standing doubts about her attractiveness, and a linked dysfunctional assumption: 'If someone is not there for me all the time, it means they don't care about me'.

(e) Background

(i) Previous treatment

Many depressed patients presenting for CBT have already received other treatment (most commonly antidepressant medication and counselling). The therapist should inquire about the **outcome** of such treatment, and **what the patient makes of it**. Depressed patients often conclude that the incomplete success of prior treatment means that they will also be unable to benefit from this new approach. Such thoughts can be worked with using straightforward cognitive behavioural methods (identify, question, test). When the patient has received psychological therapy, it is often helpful to ask **what they learned** from it. If they feel they learned nothing, this may predispose them to approach CBT with pessimism.

Alternatively, if they learned something of value, CBT can build on this positive experience.

(ii) Expectations of CBT

These may reflect general pessimism about change, especially if patients know nothing about the approach. Those who have heard or read about CBT, and are aware of outcome data, may be more optimistic, although still anxious lest it fail to help *them*. Others may have heard negative reports (e.g. that it is mechanistic and fails to address deep issues). **Non-defensive discussion** of expectations allows doubts and misconceptions to be addressed, as well as encouraging open-mindedness and continued frank feedback from the patient.

(iii) Early experience and resultant beliefs and assumptions

CBT is **conceptualization driven**,[26] **closely based on the cognitive model of depression**. In order for patients to understand how their problems developed, and that current beliefs are learned opinions rather than reflections of fact, it is helpful to know about relevant formative experiences. This is particularly true when working with severe, chronic problems, and personality disorders. However, **obtaining historical information is often not an immediate priority**. It is more important initially to convey the hopeful message that something can be done to change things for the better. Therefore details about history, beliefs, and assumptions may not be explored until work on behavioural deficits and negative automatic thoughts has produced a reduction in hopelessness and depressed mood. That said, where patients are relatively mildly depressed, psychologically minded, and able to articulate their difficulties with ease, a draft conceptualization incorporating historical information, beliefs, and assumptions sometimes emerges even from the initial assessment.

Managing treatment

(a) Starting treatment

The **first treatment interview has four main objectives**:

- to establish a warm, collaborative therapeutic **alliance**
- to list specific **problems and** associated **goals**, and select a first problem to tackle
- to educate the patient about the **cognitive model**, especially the vicious circle that maintains depression
- to give the patient first-hand experience of the focused, workman-like, empirical **style** of CBT.

These convey two important messages: (1) it is possible to make sense of depression; (2) there is something the patient can do about it. These messages directly address hopelessness and helplessness.

(i) Identifying problems and goals

The **problem list** usually includes symptoms of depression. It may also contain aspects of other disorders (e.g. panic attacks), problems in living (e.g. family conflicts), and, in some cases, long-standing psychological problems (e.g. fear of intimacy). Developing the list provides the therapist with a 'map of the territory', which suggests possible targets for intervention, as well as an opportunity to foster the therapeutic relationship by demonstrating empathy. It suggests that apparently chaotic experience can be broken down into manageable problem areas. **Goal identification** then implies that progress is possible.

(ii) Introducing the cognitive model of depression

The therapist's next task is to demonstrate **how negative thinking influences emotion and behaviour**, relating this to the patient's experience using material derived from the session. The therapist explains that the patient will learn to notice negative automatic thoughts, to stand back, question them, and develop more realistic and helpful perspectives. Patients are often doubtful about their ability to do this. It is important therefore to present **CBT as a learning opportunity** during which skills can be acquired, step by step, with the therapist's guidance. The therapist is not obliged to convince patients that CBT will work for them, but a willingness to try it in practice is essential.

(iii) Where to start?

A **first treatment target** is chosen towards the end of the first session, and an appropriate **homework** task agreed upon. Suitable homework tasks include: observing a recording of the session and noting important points (this assignment follows every session), reading assignments,[27] and self-monitoring assignments. The initial target varies. Where patients are only relatively mildly depressed, and remain active and capable of experiencing interest and pleasure, monitoring negative automatic thoughts can begin right away. Where the depression is more severe, with significant behavioural and motivational deficits, it is best to begin with behavioural interventions.

(b) Behavioural interventions

(i) Reducing rumination

In severe depression, access to more positive perspectives may be blocked by depressed mood, making modifying negative automatic thoughts difficult or impossible. At the same time, depressed patients spend a great deal of time ruminating about their difficulties and shortcomings.[28] Learning to **direct attention elsewhere** reduces the frequency of negative thoughts and hence improves mood. This palliative measure will not resolve the patient's problems, but its impact reinforces the model, and feeling somewhat better can facilitate more constructive thinking.

(ii) Monitoring activities

Lowered activity levels and loss of interest and pleasure are often central to depression, and interventions designed to address them are known to be powerful in their own right.[19] Early behavioural interventions serve to **maximize engagement in activities providing a sense of pleasure and mastery**. This has a direct impact on mood, and provides opportunities to test negative thoughts that block engagement and, in a more global sense, prevent recovery (e.g. 'I can't do anything to change how I feel').

Patients record what they do, hour by hour, on a Weekly Activity Schedule (Fig. 6.3.2.3.2). Each activity is rated **0–10 for Pleasure (P) and Mastery (M)**. P ratings indicate how enjoyable the activity was, and M ratings how much of an achievement it was. 'P' is usually easily understood, but M can present difficulties. Depressed people often feel that nothing they do is an achievement, perhaps because most of their activities are routine ('What's so special about that?') or do not meet their standards ('I should have done more'). M should therefore be explained as 'an achievement, *given how you felt at the time*'. Thus even simple activities (e.g. making a cup of tea) are real achievements when patients are hampered by low mood and loss of energy. Ratings should be made immediately

Name _____ Week beginning _____

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
9–10			Bed (P5, M0)	Take kids to school (P0, M9)	Bed (P4, M0)		
10–11			Bed (P4, M0)	Shopping with friends	Bed (P0, M0)		
11–12			Bed (P2, M0)	↓ (P6, M7)	Bed (P0, M0)		
12–1			Bed (P1, M0)	Lunch with friend (P7, M3)	Up, shower, dress (P2, M7)		
1–2			Bed (P0, M0)	Went for a walk (P7, M5)	Lunch, read paper (P3, M4)		
2–3			Up, shower, dress (P0, M5)	Gardening (P4, M7)	Therapy Session		
3–4			Pick up kids from school (P3, M7)	Pick up kids from school (P1, M3)			
4–5			Tea with friends (P4, M9)	Tea & TV with kids (P5, M2)			
5–6			Tea with friends (P6, M7)	Ironing (P0, M5)			
6–7			Get supper (P2, M5)	Get supper (P1, M5)			
7–8			Read with kids (P8, M3)	Read with kids (P7, M3)			
8–12			TV (P3, M1) Bed (P8, M0)	TV (P4, M0) Bed (P9, M0)			

Fig. 6.3.2.3.2 Weekly activity schedule. P, pleasure; M, mastery.

after each activity, since retrospective ratings may be distorted by negatively biased recall. In addition, it is helpful for patients to **review each day**, asking questions like: 'What worked for me?' 'What did not work?' 'What do I need more of? Less of? Different?'

(iii) Planning activities

Once self-monitoring is mastered, **each day is planned in advance** on an hour-by-hour basis. This:

- provides a structure and helps with setting priorities
- averts the need to keep making decisions about what to do next
- reduces what may seem like chaos to a manageable list
- increases the chances that activities will be carried out
- enhances patients' sense of control.

A pattern of activities is sought in which **mastery and pleasure are balanced and maximized**. The plan is likely to contain a blend of obligations (e.g. the ironing) and pleasures (e.g. listening to music). Avoided tasks can be included, broken down into manageable steps ('graded task assignment'). Again, it is helpful for patients to review each day in detail, identifying unhelpful thoughts to be worked on in the next therapy session (e.g. 'If I can't complete the task, I might as well not bother at all'). Thus the cognitive element is present even when behaviour change is the primary target.

(c) Working with negative automatic thoughts

Once behavioural methods have been mastered, patients learn to **identify, question, and test negative automatic thoughts**. The main tool here is the Dysfunctional Thoughts Record (Fig. 6.3.2.3.3). The example summarizes a lengthy discussion that took place when a patient experienced a serious setback midway through treatment.

(i) Identifying negative automatic thoughts

Patients learn to **record upsetting incidents** as soon as possible after they occur (delay makes it difficult to recall thoughts and feelings accurately). They learn:

1 **To identify unpleasant emotions** (e.g. despair, anger, guilt), signs that negative thinking is present. Emotions are rated for intensity on a 0 to 100 scale. These ratings (though the patient may initially find them difficult) help to make small changes in emotional state obvious when the search for alternatives to negative thoughts begins. This is important, since change is rarely all-or-nothing and small improvements may otherwise be missed.

2 **To identify the problem situation**. What was the patient doing or thinking about when the painful emotion occurred (e.g. 'waiting at the supermarket checkout', 'worrying about my husband being late home')?

3 **To identify negative automatic thoughts associated with the unpleasant emotions**. Sessions direct the therapist towards asking: 'And what went through your mind at that moment?' Patients become aware of thoughts, images, or implicit meanings that are present when emotional shifts occur, and record them word-for-word. Belief in each thought is also rated on a 0 to 100 per cent

Date	Emotion(s) *What did you feel? How bad was it (0–100%)?*	Situation *What were you doing or thinking about?*	Automatic thoughts *What exactly was running through your mind? Write your thoughts down, word for word. How far did you believe each of them (0–100%)?*	Alternative views *What alternatives are there to the automatic thoughts? How far do you believe each of them (0–100%)?*	Outcome *1. How far do you now believe the original thoughts (0–100%)? 2. How do you now feel (0–100)? 3. What can you do now (action plan, experiment)?*
Thurs	Depressed (95%) Hopeless (90%)	Three terrible days	I'm back to square one—I've lost everything I learned (100%)	Not true — even now, I'm not as bad as when I came into hospital (100%)	1. (60%)
				I am doing my housework, looking after the children, doing my job. I am getting some satisfaction out of it — it's not a total failure (75%)	2. Depressed (70%) Hopeless (35%)
			There's no point in doing anything. Nothing will work (100%)	I've been feeling very bad, but setbacks are to be expected — disappointment at the contrast with last week makes it worse (100%)	3. Accept setbacks as part of recovery, not the end of the world.
			I've tried everything now, and nothing has changed (100%)	I've been getting this therapy for 7 weeks — I've been depressed for 3 years. It's not surprising I haven't got over it completely. Already I can manage 75% of my depression, as opposed to 25% (100%)	Make a detailed plan to help me deal with these feelings if they come again.
			I've failed again (100%)	This setback is not my failure —it is part of the problem (80%)	Keep using what I have learned to deal with my depression
			I will always be like this (100%) The only solution is to kill myself (95%)	Suicide is not the answer. Keep working on your thoughts. The past 7 weeks, and today, show it can work (100%)	

Fig. 6.3.2.3.3 Dysfunctional thoughts record.

scale (100 per cent represents complete belief, 50 per cent a moderate degree of belief, and so on). Again, this helps to make small changes in conviction evident at the next stage.

The skill of identifying painful emotions and associated thoughts is best learned if therapist and patient **work through examples on the sheet** before the patient self-monitors independently. Therapists can make sure that patients understand what is required, and are prepared for possible difficulties. For example, patients sometimes avoid recording thoughts because doing so is upsetting. Therapists can reassure them that this phase will pass once they learn to answer their thoughts, and suggest that they follow recording by engaging in an absorbing and pleasurable activity. Sometimes thoughts recorded do not seem to 'fit' the emotion experienced; in this case, therapists may need to help patients to **'unpack' the meaning of the thought** (for example, 'That didn't go too well' may turn out to mean 'I'm a total failure'). Time taken to learn accurate self-monitoring varies; many patients acquire the skill within a few days, but others take much time and coaching.

(ii) Questioning negative automatic thoughts

Once patients can record thoughts and feelings, they learn to **search for alternative views**, writing these in the fifth column of the Dysfunctional Thoughts Record. There is no such thing as a 'right' answer to a negative thought; the 'right' answer is the one that helps the patient to feel better and handle the situation more constructively. Accordingly, the therapist's task is not to suggest alternatives, but rather to elicit them through '**guided discovery**', a process of sensitive questioning, which allows patients to reach new interpretations independently. It is helpful for therapists to develop a personal 'library' of questions, through discussion with colleagues, observation of other therapists, attendance at workshops, and reading. Productive areas of inquiry include:

1 **What is the evidence?** Processing biases in depression mean that patients give weight to information consistent with prevailing perspectives at the expense of information, which suggests that they may not be wholly true. The therapist thus needs to examine 'evidence' believed to support the thought, and also to seek information that might contradict it.

2 **What alternative views are there?** Questions such as the following can prompt alternative perspectives: 'How would you have reacted to this before you became depressed?' 'What is your perspective on this when you feel relatively well?' 'What might someone whose views you trusted make of this?' 'If someone you cared about came to you with this problem, what would you say?'

3 **What are the advantages and disadvantages of this way of thinking?** This approach is particularly helpful with self-critical thinking. Patients often believe that self-criticism is an effective way of bringing about change; in fact, it only intensifies depression. Patients who habitually self-criticize can be helped to draw

up an analysis of pros and cons. Apparent advantages (e.g. 'It keeps me on my toes') may in fact be outweighed by disadvantages (e.g. 'It paralyses me').

4 **What are the biases in my thinking?** The tendency to make inferential errors such as overgeneralization has already been mentioned. Learning to recognize these can be helpful, especially when patients regularly make the same mistake.

Alternatives reached by questioning negative automatic thoughts are recorded on the Dysfunctional Thoughts Record. The patient rates them for degree of belief, to ensure that they are *sufficiently* convincing (they do not require belief ratings of 100 per cent). If alternatives are not at all convincing, they will have no impact on the strength of the original automatic thoughts or associated emotions. These are now re-rated in the final column as a check that plausible alternatives have been found.

As with self-monitoring, these skills are **best learned by working through examples in session** before the patient attempts to answer thoughts independently. Even then, patients are sometimes unable to find alternatives, especially if emotion is high. This is quite normal, given that questioning one's thoughts is a complex skill. It may be helpful to leave searching for alternatives until the storm is past. Sometimes alternatives make no difference to the original thoughts or emotions. This may be because the patient has reservations about their validity ('Yes, but . . . '), which can dealt with like other negative thoughts. Alternatively, it may be that non-verbal methods (e.g. imagery work, experiential learning) are necessary to facilitate emotional change, or that the resistant thought is a more or less direct statement of an underlying belief of much longer duration, which will take longer to change.

(iii) Testing negative automatic thoughts: what can I do now?

It is important that cognitive changes brought about by questioning are consolidated through **behavioural experiments**.[17] These are often designed to test out the validity of the new perspective by seeking further information or acting differently and observing the results. They may also include practical plans to solve genuine life problems and to deal with the trigger situation differently should it occur again.

(d) Ending treatment

Although most episodes of depression are time-limited, **relapse and recurrence are common**—the more so, the more episodes a person has experienced. CBT therefore emphasizes working on cognitive vulnerability factors, summarizing and consolidating learning, and preparing for possible setbacks. A new approach, *Mindfulness-Based Cognitive Therapy*,[29] which integrates elements of CBT with intensive meditation practice, has been developed specifically to tackle this problem. Its effectiveness with patients who have experienced three or more episodes of depression has been demonstrated in two clinical outcome trials.[30,31]

(i) Re-evaluating dysfunctional assumptions

Once patients are skilled at answering negative automatic thoughts, attention turns to dysfunctional assumptions that make them vulnerable to depression. Often these emerge from information gathered earlier, for example repeating themes in Dysfunctional Thoughts Records. They may also be identified using a 'downward arrow' technique, which involves identifying situations that typically distress the patient, and associated thoughts. Instead of responding directly to these, the therapist asks: 'If that was true, what would it mean to you?' This question (or variants) is repeated until a general assumption or rule, relevant to a range of situations, emerges. The validity of the rule is then questioned and tested. This process normally takes several sessions to complete. A helpful sequence of questions is given below (these are not exhaustive:

1 **Where did this rule come from?** Identifying the source of a dysfunctional assumption (e.g. parental criticism) often helps to encourage distance by suggesting that its development is understandable, though it may no longer be relevant or useful.

2 **In what ways is the rule unrealistic?** Dysfunctional assumptions do not fit the way the world works. They operate by extremes, which are reflected in their language (always/never rather than some of the time; must/should/ought rather than want/prefer/would like).

3 **In what ways is the rule helpful?** Dysfunctional assumptions are not usually wholly negative in their effects. For example, perfectionism may lead to genuine high-quality performance. If such advantages are not recognized and taken into account when new assumptions are formulated, the patient may be reluctant to move forward.

4 **In what ways is the rule unhelpful?** The advantages of dysfunctional assumptions are normally outweighed by their costs. Perfectionism leads to rewards, but it also undermines satisfaction with achievements and stops people learning from constructive criticism.

5 **What alternative rule might be more realistic and helpful?** Once the old assumption has been undermined, it is helpful to formulate an explicit alternative (e.g. 'It is good to do things well, but I am only human-sometimes I make mistakes'). This provides a new guideline for living, rather than simply undermining the old system.

6 **What needs to be done to consolidate the new rule?** As with negative automatic thoughts, re-evaluation is best made real through experience: behavioural experiments. These encourage patients to challenge specific examples of old rules, as well as testing out the validity of new ones by acting as if they were true and observing the results This systematic work may need to continue for weeks or indeed months, given that assumptions have often been in place for many years.

(ii) Re-evaluating negative core beliefs

Negative thoughts often disappear as patients recover, whether the depression is treated by psychological means or not. Sometimes, however, they reflect enduring beliefs about the self, the world (including other people), or the future, which if left untouched may predispose the patient to become depressed again. Methods for dealing with these have primarily been developed in the context of CBT for personality disorder,[32,33] but can often be used within short-term CBT. This is important, given limited resources.

The cognitive model suggests that negative beliefs contributing to vulnerability to depression are (like dysfunctional assumptions) based on early learning, and maintained by a consistent bias in favour of information, which confirms them, and against information, which contradicts them. Therapists help patients to become aware of this bias, to question the 'evidence' that upholds

the negative beliefs (much as the 'evidence' in favour of negative automatic thoughts is questioned), and to search actively for information which contradicts it. Once a relevant belief has been identified (e.g. 'I am no good') and rated for degree of belief (0–100 per cent), the suggestion is introduced that this may be more of an opinion than a fact (work at the level of automatic thoughts should have prepared the ground for this idea). If possible, the patient is asked to suggest a more positive alternative (e.g. 'If you were not 'no good', how would you like to be?'), and belief in the alternative (which is likely to be low) is also rated. The alternative provides a new 'address' at which to store information inconsistent with the old belief. However, it is not always possible to find one at this stage (e.g. when the patient has predicated his or her life on the belief and accumulated a large body of supporting evidence). In this case, an alternative may only become available once the old belief has been systematically weakened.

Supporting 'evidence' may include events from the distant past, which have been interpreted in a self-derogatory way (e.g. childhood abuse), as well as later experiences (e.g. a broken marriage) and everyday events of the kind already recorded on the Dysfunctional Thoughts Record. Each item is questioned, and new and more adaptive interpretations arrived at. In addition, patients are asked to record evidence that would support a more positive alternative to the old belief (e.g. examples of their strengths and skills). The success of these interventions is assessed by repeatedly rating the degree of belief in the old system, as well as in the new alternatives. This work too may take a considerable time, especially if negative beliefs have had a sizeable impact on the person's life. Where treatment time is limited for practical reasons, clinicians may find it helpful to space out later sessions, ensuring that intervening weeks are used to consolidate and extend within-session work.

(iii) Consolidating learning: 'blueprints'

Preparation for ending treatment begins with the treatment contract. The implication of offering a limited number of sessions is that treatment will end, and that patients will acquire the skills necessary to deal with depression independently. Throughout therapy, they are encouraged to take increasing responsibility for determining session content, making practical suggestions, devising homework assignments, summarizing learning, and applying new skills in fresh areas. Written session summaries and therapy tapes encourage reflection and consolidation.

At the end of treatment, gains are summarized in a personal action plan or 'blueprint for the future'. The blueprint is confined to one or two sheets of paper, guides continued learning, and helps deal with relapse or recurrence. It draws on the case conceptualization, session notes, homework records, reading materials, and the like. The therapist should examine sessions, records, etc. independently, so that the plan is drawn up jointly, nothing important is forgotten, and the patient goes away with as full a summary as possible. The following questions provide a useful framework:

1 **How did my problems develop?** (unhelpful beliefs and assumptions, the experiences that led to their formation, events precipitating onset)

2 **What kept them going?** (maintenance factors)

3 **What did I learn from therapy that helped?** (techniques (e.g. activity scheduling) and ideas (e.g. 'I can do something to influence my mood'). Techniques should be detailed so that patients know exactly what to do should depression recur. Examples of handouts and record sheets can be included.)

4 **What were my most unhelpful negative thoughts and assumptions? What alternatives did I find to them?** (summarized in two columns)

5 **How can I build on what I have learned?** (a solid, practical, clearly specified action plan)

(iv) Preparation for setbacks

The blueprint should also be used to plan for relapse. It is helpful right at the beginning of treatment to tell patients that, however well they do, they may well experience a setback at some point, not least because periods of low mood are a normal part of human experience. CBT will not prevent the patient from ever having another moment's distress; it will provide tools for dealing with distress more effectively. This information can help patients to respond with less fear and despair when they do encounter setbacks.

Preparation for setbacks can be framed by the following questions:

1 **What might lead to a setback for me?** For example, future losses (e.g. children leaving home) and stresses (e.g. financial difficulties), i.e. events which impinge on patients' vulnerabilities and are thus liable to be interpreted negatively. For people who have experienced recurrent depression, mild normal low mood (without any major environmental stimulus) can act as a trigger for negative thinking which, if unchecked, can spiral down into clinical depression.

2 **What early warning signs do I need to be alert for?** Feelings, behaviours, and symptoms that might indicate the beginning of another depression are identified and listed, using careful analysis of this and previous episodes and of fluctuations in mood occurring during treatment.

3 **If I notice that I am becoming depressed again, what should I do?** Clear simple instructions, which will make sense despite low mood, are needed here. Specific ideas and techniques summarized earlier in the blueprint should be referred to. General encouragement can also be included (e.g. 'Don't panic'), as well as a specific plan for what to do if cognitive behavioural methods do not lift mood within a specified period (e.g. contact the general practitioner about medication, contact the therapist for telephone discussion or booster sessions, contact emergency services or telephone helplines in the event of serious suicidal thoughts). Recontacting therapists is often difficult, as patients may feel that they have failed or them down. Therapists should make it clear that they consider it a sign of courage to ask for further help, not a sign of weakness.

(e) Training

Cognitive behaviour therapy for depression is a sophisticated treatment, requiring theoretical knowledge, research familiarity, and clinical expertise. The latter is best developed through practical training and close supervision. Core skills have been operationalized in measures of therapist competency, such as the *Cognitive Therapy Scale*,[34] which allow practitioners to judge for themselves whether they are indeed practising cognitive behaviour therapy and to monitor skills development.

Treating depressed patients with cognitive behaviour therapy is a challenge, especially with severe, chronic, and relapsing depressions

and depression comorbid with other conditions. Therapist competency (the ability to carry out the treatment as intended and to an adequate standard) has a direct impact on outcome. Thus the need for experienced therapists, and for adequate training and ongoing supervision, especially with more difficult groups, cannot be overemphasized.[10,11]

Where established training institutes exist, it is wise to take advantage of their expertise. Even where no local institute is available, expert therapists from established centres often travel internationally to deliver conferences, workshops, seminars, and supervision—notably the triennial World Congress of Cognitive and Behaviour Therapies.

However, novice therapists will sometimes have little option but to supervise themselves, using clinical texts as a knowledge base. Therapists at all skill levels will benefit from regularly monitoring audio or video recordings of their treatment sessions, using the *Cognitive Therapy Scale* to identify strengths and areas in need of improvement. Even if no more experienced practitioner is available, peer supervision with interested colleagues is helpful. It provides external feedback on clinical practice, and makes other forms of learning possible (role play, study groups, etc.).

(f) Self-care

Depression is an infectious disease. It is easy for psychological therapists (especially if relatively inexperienced) to become contaminated by patients' hopelessness. Therefore supervision should include a focus on the therapist's own thoughts and feelings. Therapists with a substantial proportion of depressed clients should also ensure that they leaven their day by planning life-enhancing and pleasurable experiences, and should be prepared to use cognitive behavioural methods to address their own negative thoughts.

Training and supervision

The Beck Institute for Cognitive Therapy and Research in Philadelphia, Pennsylvania, runs extramural courses and training programmes (www.beckinstitute.org), as does the Oxford Cognitive Therapy Centre (www.octc.co.uk), who also offer a selection of CBT oriented booklets for patients and clinicians. The Academy of Cognitive Therapy offers information about CBT, training in CBT, certification as a cognitive therapist, and (once a member) access to a ListServe on which issues relating to theory, research, and clinical practice can be discussed with experienced cognitive therapists (www.academyofcognitivetherapy.org) The International Association of Cognitive Psychotherapy (www.cognitivetherapyassociation.org) also has a ListServe and hosts regular conferences in member countries. Different countries also have their own national organizations promoting CBT, for example, the American Association for Behavioural and Cognitive Therapies (www.aabt.org), the Australian Association for Cognitive and Behavioural Therapies (www.aabct.org), and the British Association for Behavioural and Cognitive Psychotherapies (www.babcp.com). These will be able to offer information about training opportunities.

Further information

A number of texts describe the theory and practice of CBT for depression in some detail.[1,17,21] Practical ideas can also be found in self-help texts for patients.

For the clinician:

Beck, J.S. (1995). *Cognitive therapy: basics and beyond*. Guilford Press, New York.

Fennell, M.J.V. (1989). Depression. In *Cognitive behaviour therapy for psychiatric problems: a practical guide* (eds. K. Hawton, P.M. Salkovskis, J. Kirk, and D.M. Clark). Oxford Medical Publications, Oxford.

Fennell, M. J. V., Westbrook, D., and Benneth-Levy, J. (2004). Depression. In *The Oxford guide to behavioural experiments in cognitive therapy* (eds. J. Benneth-Levy, G. Butler, M.J.V. Fennell, *et al.*). Oxford University Press, Oxford.

For the patient:

Beck, A.T. and Greenberg, R.L. (1974). *Coping with depression*. Available from The Beck Institute of CBT and Research, GSB Building, City Line and Belmont Avenues, Suite 700, Bala Cynwyd, Philadelphia, PA 19004–1610, USA.

Burns, D.D. (1980). *Feeling good*. New American Library, New York.

Butler, G. and Hope, A. (2007). *Manage your mind* (2nd edn). Oxford University Press, Oxford.

Fennell, M.J.V. (1999). *Overcoming low self-esteem*. Constable Robinson, London.

Gilbert, P. (1996). *Overcoming depression*. Constable Robinson, London.

Greenberger, D. and Padesky, C.A. (1995). *Mind over mood*, Chap. 10. Guilford, New York.

Westbrook, D. (1999). *Coping with depression*. Available from: Oxford Cognitive Therapy Centre, Dept of Clinical Psychology, Warneford Hospital, Oxford OX3 7JX, UK.

Subjective experience:

Solomon, A. (2002). *The noonday demon*. Simon & Schuster, New York.

Styron, W. (1991). *Darkness visible*. Jonathan Cape, London.

Wolpert, L. (1999). *Malignant sadness*. Faber, London.

Current research and theoretical developments are best followed through professional journals (primarily psychiatry and clinical psychology—see the reference list for examples), available in libraries and through the Internet.

References

1. Beck, A.T., Rush, A.J., Shaw, B.F., *et al.* (1979). *Cognitive therapy of depression*. Guilford Press, New York.
2. Beck, A.T. (1963). Thinking and depression: 1. Idiosyncratic content and cognitive distortions. *Archives of General Psychiatry*, **9**, 324–33.
3. Beck, A.T. (1964). Thinking and depression: 2. Theory and therapy. *Archives of General Psychiatry*, **10**, 561–71.
4. Beck, A.T. (1967). *Depression: clinical, experimental and theoretical aspects*. Harper and Row, New York.
5. Rush, A.J., Beck, A.T., Kovacs, M., *et al.* (1977). Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognitive Therapy and Research*, **1**, 17–37.
6. Hollon, S.D., Thase, M.E., and Markowitz, J.C. (2002). Treatment & prevention of depression. *Psychological Science in the Public Interest*, **3**, 39–77.
7. Paykel, E.S., Scott, J., Cornwall, P.L., *et al.* (2005). Duration of relapse prevention after cognitive therapy in residual depression: follow-up of controlled trial. *Psychological Medicine*, **35**, 59–68.
8. Jarrett, R.B., Kraft, D., Doyle, J., *et al.* (2001). Preventing recurrent depression using cognitive therapy with and without a continuation phase. *Archives of General Psychiatry*, **58**, 381–8.
9. Vos, T., Corry, J., Haby, M.M., *et al.* (2005). Cost-effectiveness of cognitive-behavioural therapy and drug interventions for major depression. *The Australian and New Zealand Journal of Psychiatry*, **39**, 683–92.

10. DeRubeis, R.J., Hollon, S.D., Amsterdam, J.D., *et al.* (2005). Cognitive therapy vs. medications in the treatment of moderate to severe depression. *Archives of General Psychiatry*, **62**, 409–16.
11. Hollon, S.D., DeRubeis, R.J., Shelton, R.C., *et al.* (2005). Prevention of relapse following cognitive therapy vs. medications in moderate to severe depression. *Archives of General Psychiatry*, **62**, 417–22.
12. Ingram, R.E. (2003). Origins of cognitive vulnerability to depression. *Cognitive Therapy & Research*, **27**, 77–88.
13. McGinn, L.K., Cukor, D., and Sanderson, W. (2005). The relationship between parenting style, cognitive style and anxiety and depression: does increased early adversity influence symptom severity through the mediating role of cognitive style? *Cognitive Therapy & Research*, **29**, 219–42.
14. Alloy, L.B., Abramson, L.Y., Whitehouse, W.G., *et al.* (2006). Prospective incidence of first onsets and recurrences of depression in individuals at high and low cognitive risk for depression. *Journal of Abnormal Psychology*, 115, 145–56.
15. Segal, Z.V., Gemar, M., and Williams, S. (1999). Differential cognitive response to a mood challenge following successful cognitive therapy or pharmacotherapy for unipolar depression. *Journal of Abnormal Psychology*, **108**, 3–10.
16. Lau, M.A., Segal, Z.V., and Williams, J.M.G. (2004). Teasdale's differential activation hypothesis: implications for mechanisms of depressive relapse and suicidal behaviour. *Behaviour Research and Therapy*, **42**, 1001–17.
17. Fennell, M.J.V., Westbrook, D., and Bennett-Levy, J. (2004). Depression. In *The Oxford guide to behavioural experiments in cognitive therapy* (eds. J. Bennett-Levy, G. Butler, M.J.V. Fennell, *et al.*). Oxford University Press, Oxford.
18. Tang, T.Z., DeRubeis, R.J., Beberman, R., *et al.* (2005). Cognitive changes, critical sessions, and sudden gains in cognitive-behavioural therapy for depression. *Journal of Consulting and Clinical Psychology*, **73**, 168–72.
19. Dimidjian, S., Hollon, S.D., Dobson, K.S., *et al.* (2006). Randomized trial of behavioural activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, **74**, 658–70.
20. Scott, J., Payken, E., Morris, R., *et al.* (2006). Cognitive behavioural therapy for severe and recurrent bipolar disorders. Randomised controlled trial. *The British Journal of Psychiatry*, **188**, 313–20.
21. Moore, R.G. and Garland, A. (2003). *Cognitive therapy for chronic and persistent depression*. Wiley, Chichester, UK.
22. American Psychiatric Association. (1994). *Diagnostic and statistical classification of diseases and related health problems* (4th edn). American Psychiatric Association, Washington, DC.
23. Beck, A.T., Ward, C.H., Mendelson, M., *et al.* (1961). An inventory for measuring depression. *Archives of General Psychiatry*, **4**, 561–71.
24. Beck, A.T., Weissman, A., Lester, D., *et al.* (1974). The measurement of pessimism: the Hopelessness Scale. *Journal of Consulting and Clinical Psychology*, **42**, 861–5.
25. Beck, A.T. (1993). *Beck scale for suicide ideation* (*manual*). Psychological Corporation, San Antonio.
26. Persons, J.B. (1989). *Cognitive therapy in practice: a case formulation approach*. Norton, New York.
27. Beck, A.T. and Greenberg, R.L. (1974). *Coping with depression*. Available from The Beck Institute of Cognitive Therapy and Research, GSB Building, City Line and Belmont Avenues, Suite 700, Bala Cynwyd, Philadelphia, PA 19004–1610, USA.
28. Papageorgiou, C. and Wells, A. (2003). *Depressive rumination: nature, theory and treatment*. Wiley, Chichester.
29. Segal, Z.V., Williams, J.M.G., and Teasdale, J.D. (2002). *Mindfulness-based cognitive therapy for depression: a new approach to preventing relapse*. Guilford, New York.
30. Teasdale, J.D., Segal, Z.V., Williams, J.M.G., *et al.* (2000). Reducing risk of recurrence of major depression using Mindfulness-Based Cognitive Therapy. *Journal of Consulting and Clinical Psychology*, **68**, 615–23.
31. Ma, J. and Teasdale, J.D. (2004). Mindfulness-based cognitive therapy for depression: replication and exploration of differential relapse prevention effects. *Journal of Consulting and Clinical Psychology*, **72**, 31–40.
32. Padesky, C.A. (1994). Schema change processes in cognitive therapy. *Clinical Psychology and Psychotherapy*, **1**, 267–78.
33. Beck, A.T., Freeman, A., Davis, D.D., *et al.* (2004). *Cognitive therapy of personality disorders*. Guilford Press, New York.
34. Young, J.E. and Beck, A.T. (1980). *Development of an instrument for rating cognitive therapy: the cognitive therapy scale*. University of Pennsylvania, Philadelphia, PA.

6.3.2.4 Cognitive behaviour therapy for schizophrenia

Max Birchwood and Elizabeth Spencer

Introduction

Cognitive behaviour therapy (**CBT**) for schizophrenia focuses on the core psychotic symptoms of hallucinations and delusions. Other psychosocial approaches to psychosis (e.g. intervention with families and to promote medication compliance) also frequently use CBT techniques. In this chapter, however, we focus on CBT for delusional beliefs and other psychotic phenomena and review evidence for its efficacy.

Background: assumptions and common components

The CBT approach to psychotic symptoms comprises two different strands each with their own theoretical basis, although of late these two approaches have become conjoined in practice.

Coping strategy enhancement

The first approach is inspired by the stress-vulnerability model of schizophrenia. Vulnerability here is viewed as a 'black box', drawing mainly on the biomedical tradition. It is assumed that stressors capable of triggering or exacerbating symptoms may be generated or modulated by the individual. For example, stressors emanating from the social environment are modulated by the patient's own appraisal of their stressfulness and his or her coping strategies.

Another class of stressors consists of the symptoms themselves. It is assumed that certain strategies used to cope with symptoms are unhelpful and generate stress in the individual, in turn, exacerbating symptoms. These strategies are conventionally divided into affective strategies (e.g. relaxation, sleep, etc.), behavioural strategies (being active, drinking alcohol, etc.), and cognitive strategies (distraction, challenging voices, switching attention away from voices, etc.). This underpins the approach known as Coping Strategy Enhancement[1] whereby patients are offered a range of strategies which are implemented in an empirical fashion to determine their effectiveness in symptom control. For example, Falloon and Talbot[2] documenting the coping strategies used by voice-hearers, concluded that those who had multiple strategies available

to them were more able to cope with their voices. Tarrier[3] on the other hand, focusing on a wider range of psychotic symptoms, concluded that those who applied strategies consistently tended to fare the best. This approach views the individual as an active agent who attempts to reduce the threat or distress posed by psychotic symptoms, but does not concern itself with the content or meaning that psychotic symptoms may have to the individual. There is also, in this approach, assumed to be a fundamental discontinuity between normal and abnormal functioning that comes about once the biological vulnerability is 'online'.

CBT for delusions and hallucinations

The second CBT strand draws its theoretical strength from the cognitive therapy approach.[4] Early work in this area focused on the similarity between normal (but strongly held) beliefs and delusions, in terms of the psychological processes at play in their maintenance. For example, Brett-Jones *et al.*[5] showed that delusions, like everyday beliefs, lead the individual to recruit evidence to support them and to de-emphasize or dismiss contradictory evidence. Continuing the exploration of the continuity between normal and delusional beliefs, Birchwood and Chadwick[6] argued that certain beliefs about voices' power may be considered as a quasi-rational response to anomalous experience, with the meaning attributed to them in terms of identity, power, and the consequence of disobedience determining distress and behaviour in relation to the voice. Other work has drawn on the cognitive therapy approach in depression, which emphasizes the importance of evaluative beliefs about the self (e.g. self-worth) in the genesis and maintenance of depressed mood.[7] The application of this to psychosis also emphasizes evaluative beliefs about the self. The precise relationship between self-evaluative beliefs and delusional thinking is a much debated issue of the present time. It has been argued, for example, that delusions may serve the function of defending the individual from the full impact of low self-worth through blaming others for negative events rather than the self. This is the so-called 'paranoid defence'.[8] The content of psychotic thinking often reflects such personal issues. For example, for the patient who has been sexually abused, this theme tends to crop up in the content of voice activity or in the supposed identity of the voice.[6]

This early work has been elegantly drawn together in a cognitive model of psychosis by Phillipa Garety and her colleagues. In a seminal paper[9] they propose a model of the cognitive processes leading to the positive symptoms of psychosis. In brief, positive symptoms in psychosis are hypothesized to begin with basic cognitive disturbances with lead to ambiguous sensory input, the intrusion into consciousness of unintended material from memory, or to difficulties with the self-monitoring of intentions and actions, such that they are experienced as alien. This result in anomalous conscious experiences such as actions being experienced as unintended, racing thoughts, thoughts appearing to be broadcast, and thoughts experienced as voices.

However, the authors argue that such anomalous experiences alone do not develop into full-blown psychotic experiences unless an individual appraises them as externally caused and personally significant. Such appraisals are the results of particular reasoning processes (e.g. data gathering bias or externalizing attributional style), dysfunctional personal schemas (e.g. low self-esteem born of adverse social experience), emotional states (e.g. anxiety and depression), and appraisal of the experience of illness.

This model integrates the two strands of cognitive therapy. It suggests, for example, that the reduction in dysfunctional emotional states through, for example, coping strategy enhancement, will contribute to alterations in the attributions, which are important in the formation and maintenance of the positive symptoms. Similarly, it provides a theoretical basis for the basic techniques traditionally used in the second strand of CBT. Such techniques encourage the individual to weigh evidence that contradicts a delusion as a strategy to compensate for the basic information-processing abnormality, challenge negative self-schemata, and combat depression.

Evaluation

In recent years the volume of trials evaluating CBT for psychosis has greatly expanded, with approximately 20 such studies now reported. Most of these have been conducted in the United Kingdom among patients with chronic schizophrenia. The strength of the data is now sufficient for The UK National Institute for Clinical excellence to state that cognitive behavioural therapy should be available as a treatment option for people with schizophrenia.[10]

An up-to-date review[11] reporting the analysis of 19 studies of CBT for positive symptoms in schizophrenia, found a mean effect size of 0.37. The authors concluded that 74 per cent of these studies achieved small effect sizes, 32 per cent moderate effect sizes, and 16 per cent large effect sizes in improving positive psychotic symptoms, relative to standard psychiatric care. Furthermore, they argued that this is unlikely to be due to publication bias. The effect sizes for CBT versus standard treatment among patients with chronic illness were greater than those among acutely ill patients. This may have been due to a ceiling effect caused by the effectiveness of medication in reducing symptoms in floridly ill inpatients. Similarly, the better the design of the trial, the smaller the treatment effect size, suggesting that CBT is not the panacea for all psychotic ills that it may have originally appeared.

With regard to relapse prevention, CBT appears to be more successful when the intervention is focused on relapse prevention, rather than relapse prevention being one of a series of components.[11] For example, Gumley and his colleagues[12] were able to demonstrate that a group of patients with psychosis receiving targeted CBT for relapse prevention had almost half the rate of relapse over a 12-month period compared with a similar group receiving treatment as usual (18.1 versus 34.7 per cent). In this study, the CBT treatment consisted of an engagement phase, early signs of relapse monitoring with a personalized questionnaire, and targeted CBT at the first sign of impending relapse.

While large, pragmatic trials of CBT treatment packages have yielded the above favourable results, investigating the active elements of the interventions is difficult because CBT for psychosis now refers to a wide range of treatments.

Furthermore, although the conceptual basis of CBT emphasizes the link between emotion, cognition, and behaviour, modifying emotion in psychosis has been relatively neglected in CBT trials, in favour of outcomes based on modification of delusions and hallucinations themselves.

It has been proposed that CBT should be focused into more targeted interventions aimed at emotional dysfunction or distress and/or behavioural anomaly in psychosis that is directly or indirectly linked to psychosis symptoms.[13] This approach recognizes

that while changing the psychosis symptoms might not always be possible, it may well be feasible to change the affective consequences of the symptoms or the diagnosis. This affective change may have further benefits in reducing the severity of the psychosis experience *per se*.

For example:

1 CBT can be used to reduce distress, depression, and problem behaviour associated with commanding voices, without changing the frequency or content of the voices themselves.[14]

2 CBT can focus on anxiety, depression, and interpersonal difficulty in individuals at high risk of developing psychosis.[15]

3 CBT can focus on the relapse prodrome to prevent relapse in psychosis.[12]

4 CBT can focus on 'comorbid' depression and social anxiety, including the patient's appraisal of the diagnosis and its stigmatizing consequences.[16]

5 CBT can be used to reduce stress reactivity, thereby increasing resilience to life stress and preventing psychotic relapse.[17]

6 CBT can be used to increase self-esteem and social confidence in people with psychosis.[18]

Management

Coping strategy enhancement

Coping strategy enhancement involves developing a coping repertoire and over-rehearsing it to facilitate an automatic coping response.[19] It can either be used to improve an individual's attempts to cope with his or her voices by developing an understanding of factors that trigger or improve the voices, or to test the reality of thoughts about the voices.

For example, therapy with a patient who hears threatening and frightening voices exacerbated of being alone might involve: An explicit congratulation on the strength involved in withstanding the voices' incessant activity; strengthening of coping strategies involving seeking company and social support; and the use of a personal stereo for distraction.

Cognitive therapy to challenge delusions and dysfunctional assumptions

The application of cognitive therapy in challenging of delusions and dysfunctional beliefs draws upon the approach described by Chadwick *et al.*[20] and builds upon the pioneering work of Chadwick and Lowe.[21]

Engaging patients is perhaps the greatest challenge facing a therapist. It is noticeable that many individuals either never attend or do so for a few sessions and then stop. Once individuals get past the opening strategies of cognitive therapy they usually see therapy through. Careful attention to appropriate therapeutic technique can maximize client engagement. Similarly, Kuipers *et al.*[22] report that a response to CBT in their study was associated with greater cognitive flexibility concerning delusions at baseline. This suggests that the sufferer may need some small degree of insight into the fact that he might be mistaken, to benefit from cognitive therapy for delusions.

The process involves six basic steps as summarized in Table 6.3.2.4.1.

Table 6.3.2.4.1 A summary of the steps in cognitive therapy for delusions

1 Viewing delusions as beliefs, not facts
2 Developing a rationale for questioning the delusion
3 Weakening delusions
4 Utilizing inconsistency and irrationality
5 Reformulating delusions as reactions to, and attempts to make sense of, specific experience
6 Assessing the delusion and alternative, and empirical testing

(a) Viewing delusions as beliefs, not facts

The first technical difficulty to be encountered is the necessary move to aid the client in conceptualizing a delusion as a belief and not a fact. This move is an essential part of CBT for all emotional problems, but is difficult at the best of times. With depression, for example, patients often struggle to appreciate that their sense of worthlessness, which is so concrete to them, is actually a belief they hold and is different from a knowledge of events and facts. With delusions there is the added complication that the therapist might be perceived as being just another person who disbelieves the patient.

There are two central points to bear in mind when seeking to reconceptualize delusions as beliefs, not facts—why it is being done and how it is done.[20]

The purpose of clarifying that delusions are beliefs, not facts, is to empower the patient and offer a way of easing his or her distress. If the patient really is being persecuted by a powerful organization, or has a radio transmitter and receiver in his head, neither he nor the therapist can actually change this. The patient feels that he knows this as a fact, with the consequence that he feels frustrated and helpless as well as distressed. However, if the patient only believes this to be true, then he gains the freedom to examine his beliefs and perhaps change his distressing feelings and behaviour and experiences himself. In this sense it is in his best interest for the delusion to be false.

How this process takes place is critical. The process of Socratic questioning is not one of persuading a patient that he is wrong and that you, the therapist, are right. This mistake is made all too often. Rather, in Socratic dialogue the therapist helps the patient to draw on his own doubt and experience in order to realize that there are other ways in which he is able to make sense of his experience. So, when the therapist pursues the conceptual step of clarifying that a delusion is only a belief, the patient's own doubt, past or present, his own contradictory experience and behaviour, and concern about the possibility that the delusion is wrong are accessed. Many patients have 'double awareness' of delusions—on the one hand they believe them firmly and are distressed and disturbed by them, yet on the other hand they behave in ways that contradict the delusion, and they believe that working with a therapist might ease the problem. Finally, the therapist must accept that it is acceptable if the patient does not alter his belief. The process is 'collaborative empiricism, not indoctrination'.[17]

(b) Developing a rationale for questioning the delusion

Patients are usually used to being told by family and carers that their beliefs are wrong, that they are deluded. It is easy for a therapist to

prepare the intervention well and embark on it before the patient is clear of the purpose and possible benefit, thus causing early loss of engagement. It is revealing to turn the engagement question on its head and to consider why a patient should ever wish to engage in therapy. With emotional problems patients identify their problems as depression, anger, anxiety, guilt, etc.; with delusions and hallucinations this is not so—patients predominantly present problems which they believe are actual events (persecution, voices, passivity). This means that they have no clear objective and therefore have no particular motivation to engage. The key reason for a patient to reconsider delusional beliefs is that it will help him feel less distress and it will free him to behave differently and to pursue the things he wants more directly. What the therapist does gradually through the unfolding cognitive assessment is to clarify with the patient that he is experiencing emotional and behavioural problems, and that these are tied to his beliefs (delusional and evaluative). The therapist then needs to explore with the patient how the delusion affects his life and how his life would be different (i.e. better or worse) if the delusion were false.[20] In this way, the therapist slowly encourages the patient to view the delusion not as an important discovery but as a belief that results in distress (e.g. fear, anxiety) and causes him to behave in ways which he would rather not (e.g. avoid things he would otherwise like to do).

(c) Weakening delusions

Disputing comprises four elements:[20]

- The evidence for the belief is challenged, in inverse order of its importance to the delusion.
- The internal consistency and plausibility of the delusional system is questioned.
- Following Maher,[23] the delusion is reformulated as being an understandable response to, and way of making sense of, specific experience, and a personally meaningful alternative is then constructed.
- The individual's delusion and the alternative are assessed in the light of the available information.
- Challenging the evidence for the belief.

Watts *et al.*[24] argued that a danger when trying to modify delusions, indeed, all strongly held beliefs, was psychological reactance, whereby too direct an approach served only to reinforce the belief. They offered two principles to minimize this possibility: begin with the least important belief, and also work with the evidence for the belief rather than the belief itself.

Accordingly a 'verbal challenge' of delusions begins by questioning the evidence for the belief, and this process begins with the least significant item of evidence and works up to the most significant one. Our preferred approach is that with each item of evidence the therapist questions the patient's delusional interpretation and puts forward a more reasonable and probable one. The customary approach in CBT is for the patient to be asked to generate the alternative interpretation(s), rather than the therapist supply one, but we have found that for certain patients this conventional tactic is a weak intervention.

When the therapist questions the evidence for a delusion there are two distinct but related objectives. One is to encourage the patient to question and perhaps even to reject the evidence for his or her belief, and in this way perhaps to undermine the patient's conviction in the delusion itself. For some individuals challenging the evidence is a very powerful intervention and one that produces a substantial reduction in delusional conviction. However, more commonly this does not happen, but challenging evidence is still valuable in that it does impart insight into the connection between events, beliefs, affect, and behaviour. This is the second objective of challenging evidence, namely to convey the essentials of the ABC approach, i.e. that strongly held beliefs influence affect, behaviour, and cognition (i.e. interpretation) for all people. Core beliefs recruit or bias everyday inferences and automatic thoughts. However, this means that people often impose an interpretation on to events, which is unwarranted, and because we are prone towards selectively processing information that confirms our beliefs, this goes undetected. In other words, it is understandable that a patient should interpret a particular event in line with his delusion because this is merely one occurrence of a general tendency, and confirmation bias, common to all of us. In therapy, it is helpful to convey the ordinariness and normality of this process with everyday examples.

Having considered the alternatives, the patient is then asked to rate his conviction about each; regardless of how convinced he remains that the delusional interpretation is correct, it is usual to move on to the next piece of evidence. The therapist does not have to change what the patient thinks, but only to offer a fresh insight into the way he is thinking.

(d) Utilizing inconsistency and irrationality

Although delusions contain differing degrees of inconsistency and irrationality, they all seem to contain some. For example, Margaret believed that she could not act or make a decision without reference to her voice; however, she described periods when she was relaxed and the voices quiescent where she would be making decisions. Such inconsistencies can be therapeutically useful.

(e) Reformulating delusions as reactions to, and attempts to make sense of, specific experience

We always construe a delusion as both a reaction to and an attempt to make sense of certain puzzling and often threatening experience. It is an understandable and reasonable attempt to find meaning at a time when the individual is bewildered, anxious, and frightened. But, the delusion carries a cost in terms of distress and disturbance, which the individual might not otherwise experience. This is how delusions may be explained to patients.[23]

At this stage the therapist has commenced the process of challenging the delusional belief, re-formed the delusion as an attempt to make sense of certain experiences (e.g. primary symptoms, trauma) and raised the idea that the delusion is psychologically functional (i.e. it eases puzzlement) and perhaps linked to evaluative beliefs.

(f) Assessing the delusion and alternative, and empirical testing

Finally, the patient and therapist need to assess the delusion and alternative in the light of the available evidence and previous discussion. The therapist may spell out the advantages of the alternative interpretative framework, which can also be discussed by relating it to the patient's experiences.

It is an integral part of CBT that the belief or assumption under consideration be tested empirically. Such reality testing involves planning and performing an activity that validates or invalidates a belief, or part of a belief.

When working with delusions, we set up a clear alternative belief in opposition to the delusion, clarifying with the patient in advance precisely what has to happen for each to be supported and refuted. For example,[20] Alison believed that by repeating her voice's command (e.g. 'The price of milk *will* rise a billion times') it would actually happen by transmitting the thought to a member of the government who would act upon it. The empirical test involved repeating the voice and purchasing milk before and after doing so, predicting that within 2 weeks the price of milk would at least double. If it did not, the alternative (that the power to change events was very weak) would be strengthened.

Cognitive therapy for beliefs about auditory hallucinations

The above techniques can be applied to challenge beliefs about 'voices' using the following three steps:

(a) Assessment

Assessing the personal meaning a voice has for a person is the defining feature of the cognitive approach to assessing auditory hallucinations. The delusional beliefs found to be most significant are those relating to a voice's identity, purpose, power, knowledge, and the consequences of compliance and resistance. The semi-structured interview schedule developed by Chadwick and Birchwood[6,25] is recommended.

(b) Disputing beliefs about voices

The thrust of the therapist's challenge is that the beliefs are reasonable and understandable reactions to, and attempts to make sense of, the auditory hallucinations. The therapist reviews evidence and inconsistency, and plans tests, with the aim always of evaluating two possible meanings: that the beliefs are true, a discovery, or that they are reasonable and understandable, but mistaken. As ever, it is vital that the therapist really practises Socratic questioning and works collaboratively. This involves drawing out patients' own doubts, puzzlement, double awareness, critical faculty, etc., rather than forcing a contradiction on them.

The major piece of evidence for the delusional beliefs is always the actual voices, especially their content—these are, after all, the activating events that the delusions are invoked to explain. The role of beliefs is critical because individuals usually attribute voices with a power and knowledge that goes well beyond what they have actually said. Several examples of challenging follow.

It is really quite common for beliefs about compliance (e.g. 'If I don't do what my voice says I will be punished') not to fit patients' experiences, and perhaps it is only their emotional impact which prevents patients from abandoning them. Kate, for example, believes 'If I drop my guard the voices will kill me', but in fact she has dropped her guard on many occasions without consequence. This might be pointed out as follows. 'Kate, you say that the voices have the power to kill you and you must be on your guard constantly. I certainly appreciate the fear that this must create. What puzzles me though is that your guard is often down, like when you are asleep. How is it that they have not succeeded in all these years?'

The appearance of being all knowing (omniscience) is a vital aspect of many voices[22] and often features as a key piece of evidence that the identity of the voice is superhuman. It leaves individuals feeling exposed and vulnerable and very prone to guilt and shame. Alice believes that her voice is a prophet endowed with the ability to foretell the future. In particular, the voice anticipates exactly the arrival of her husband home from work each day. To begin the process of questioning she was asked: 'Let's suppose for a moment that the voice cannot foretell the future; can you think of any other possible explanations for last night's prediction?' One such possibility was that the voice was making a very safe guess.

(c) Testing beliefs about control

A useful strategy is to use a procedure whereby the patient and therapist learn to engineer situations to start or increase the probability of hearing voices, and then to stop or reduce them. In this way the patient gains a surprising degree of control over the voice. The initial assessment provides information about cues that provoke voices for a particular individual; concurrent verbalization is known to stop or diminish voices temporarily. This information is combined in the following five steps.

- Identify cues that increase and decrease voices.
- Practise the use of 'increasing' and 'decreasing' strategies within a session.
- Propose the notion that 'control' requires the demonstration that voice activity can be turned up/on or down/off.
- In sessions encourage the patient to initiate or increase voice activity for short periods then reduce or stop it.
- Elicit changes in the patient's belief about his control over the voices.

The above process has been applied in a targeted way to the special case of beliefs about command hallucinations. These are high-risk, distressing, and relatively common symptoms of schizophrenia.

For example, Byrne and her colleagues[26] have developed a specific cognitive therapy for command hallucinations, which draws on the above techniques. Using the methods of collaborative empiricism and Socratic dialogue, the therapist seeks to engage the client to question, challenge, and undermine the power beliefs, then to use behavioural tests to help the client gain disconfirming evidence against the beliefs. These strategies are also used to build clients' alternative beliefs in their own power and status, and finally, where appropriate, to explore the origins of the schema so clients have an explanation for why they developed those beliefs about the voice in the first place. They were able to show that this process produce significant reductions in compliance behaviours and favourable changes in beliefs about the power, superiority and need to comply with the voices, despite the frequency, loudness, and content of the voices staying the same.[14]

Conclusion

CBT for psychosis is a rapidly developing field, and one that has borne considerable fruit in terms of providing effective treatments and a basis upon which a dialogue between the patient and the professional can take place about matters of great concern and a source of much distress.

Further information

Nelson, H. (2005). *Cognitive behavioural therapy with delusions and hallucinations: a practice manual* (2nd edn). Nelson Thornes, Cheltenham.

Morrison, A.P. (2002). *A casebook of cognitive therapy for psychosis*. Brunner/Routledge, London.

Byrne, S., Meaden, A., Trower, P., *et al.* (2006). *A casebook of cognitive behaviour therapy for command hallucinations: a social rank theory*. Routledge, London.

References

1. Tarrier, N., Yusupoff, L., Kinney, C., *et al.* (1998). Randomised controlled trial of intensive cognitive behavioural therapy for patients with chronic schizophrenia. *British Medical Journal*, **317**, 303–7.
2. Falloon, I.R.H. and Talbot, R. (1981). Persistent auditory hallucinations: coping mechanisms and implications for management. *Psychological Medicine*, **11**, 329–39.
3. Tarrier, N. (1987). An investigation of residual positive symptoms in discharged schizophrenic patients. *British Journal of Clinical Psychology*, **26**, 141–3.
4. Beck, A.T., Rush, A., Shaw, B., *et al.* (1979). *Cognitive therapy of depression*. Guilford, New York.
5. Brett-Jones, J., Garety, P., and Hemsley, D.R. (1987). Measuring delusional experiences: a method and its application. *British Journal of Clinical Psychology*, **26**, 257–65.
6. Birchwood, M. and Chadwick, P.D. (1997). The omnipotence of voices. II. Testing the validity of the cognitive model. *Psychological Medicine*, **27**, 1345–53.
7. Garety, P.A., Kuipers, L., Fowler, D., *et al.* (1994). Cognitive behavioural therapy for drug resistant psychosis. *The British Journal of Medical Psychology*, **67**, 259–71.
8. Bentall, R.P. and Kinderman, P. (1998). Psychological processes and delusional beliefs: implications for the treatment of paranoid states. In *Outcome and innovation in psychological treatment of schizophrenia* (eds. T. Wykes, N. Tarrier, and S. Lewis), pp. 119–44. Wiley, Chichester.
9. Garety, P.A., Kuipers, A.E., Fowler, D., *et al.* (2001). A cognitive model of the positive symptoms of psychosis. *Psychological Medicine*, **31**, 189–95.
10. National Collaborating Centre for Mental Health. (2003). *Schizophrenia: full national clinical guideline on core interventions in primary and secondary care*. Royal College of Psychiatrists and the British Psychological Society, London.
11. Tarrier, N. and Wykes, T. (2004). Is there evidence that cognitive behavioural therapy is an effective treatment for schizophrenia? A cautious or cautionary tale. *Behaviour Research and Therapy*, **42**, 1377–401.
12. Gumley, A., O'Grady, M., McNay, L., *et al.* (2003). Early intervention for relapse in schizophrenia: results of a 12-month randomized controlled trial of behavioural therapy. *Psychological Medicine*, **33**, 419–31.
13 Birchwood, M. and Trower, P. (2006). The future of cognitive behavioural therapy for psychosis: not a quasi-neuroleptic. *British Journal of Psychiatry*, **188**, 107–8.
14. Trower, P., Birchwood, M., and Meaden, A. (2004). Cognitive therapy for command hallucinations: randomised controlled trial. *British Journal of Psychiatry*, **184**, 312–20.
15. Morrison, A.P., French, P., Walford, L., *et al.* (2004). Cognitive therapy for the prevention of psychosis in people at ultra-high risk: randomised controlled trial. *British Journal of Psychiatry*, **185**, 291–7.
16. Iqbal, Z., Birchwood, M., and Chadwick, P. (2000). Cognitive approach to depression and suicidal thinking in psychosis. 2. Testing the validity of a social ranking model. *British Journal of Psychiatry*, **177**, 522–8.
17. Myin-Germeys, I., Delespaul, P., and van Os, J. (2005). Behavioural sensitization to daily life stress in psychosis. *Psychological Medicine*, **35**, 733–41.
18. Hall, P.L. and Tarrier, N. (2003). The cognitive behavioural treatment of low self esteem in psychotic patients: a pilot study. *Behaviour Research Therapy*, **41**, 317–32.
19. Tarrier, N. (1992). Management and modification of residual psychotic symptoms. In *Innovations in the psychological management of schizophrenia* (eds. M. Birchwood and N. Tarrier), pp. 38–72. Wiley, Chichester.
20. Chadwick, P.D., Birchwood, M., and Trower, P. (1996). *Cognitive therapy for delusions, voices and paranoia*. Wiley, Chichester.
21. Chadwick, P.D.J. and Lowe, C.F. (1990). Measurement and modification of delusional beliefs. *Journal of Consulting and Clinical Psychology*, **58**, 225–32.
22. Kuipers, E., Garety, P., Fowler, D., *et al.* (1997). London-East Anglia randomised controlled trial of cognitive behavioural therapy for psychosis. I. Effects of the treatment phase. *The British Journal of Psychiatry*, **171**, 319–27.
23. Maher, B.A. (1988). Anomalous experience and delusional thinking: the logic of explanation. In *Delusional beliefs* (eds. F. Oltmans and B.A. Maher), pp. 5–22. Wiley, New York.
24. Watts, F., Powell, E., and Austin, S.V. (1973). The modification of abnormal beliefs. *The British Journal of Medical Psychology*, **46**, 359–63.
25. Chadwick, D. and Birchwood, M. (1994). The omnipotence of voices. A cognitive approach to auditory hallucinations. *British Journal of Psychiatry*, **164**, 190–201.
26. Byrne, S., Meaden, A., Trower, P., *et al.* (2005). *A casebook of cognitive behaviour therapy for command hallucinations: a social rank theory approach*. Routledge, London.

6.3.3 Interpersonal psychotherapy for depression and other disorders

Carlos Blanco, John C. Markowitz, and Myrna M. Weissman

Introduction

Interpersonal psychotherapy (IPT) is a time-limited, diagnosis-focused therapy. IPT was defined in a manual. Research has established its efficacy as an acute and chronic treatment for patients with major depressive disorder (MDD) of all ages, as an acute treatment for bulimia nervosa, and as adjunct maintenance treatment for bipolar disorder.[1–9] The research findings have led to its inclusion in treatment guidelines and increasing dissemination into clinical practice.

Demonstration of efficacy in research trials for patients with major depressive episodes (MDEs) has led to its adaptation and testing for other mood and non-mood disorders. This has included modification for adolescent and geriatric depressed patients[10,11] patients with bipolar[12] and dysthymic disorders;[13,14] depressed HIV-positive[15] and depressed pregnant and postpartum patients;[16,17] depressed primary care patients;[18] and as a maintenance treatment to prevent relapse of the depression.[5] Most of the modifications have been relatively minor and have retained the general principles and techniques of IPT for major depression.

Non-mood targets have included anorexia, bulimia, substance abuse, borderline personality disorder, and several anxiety disorders. In general, outcome studies of IPT have suggested its promise for most psychiatric diagnoses in which it has been studied, with the exceptions of anorexia, dysthymic disorder, and substance use disorders.[14,19,20]

IPT has two complementary basic premises. First, depression is a medical illness, which is treatable and not the patient's fault. Second, depression does not occur in a vacuum, but rather is influenced by and itself affects the patient's psychosocial environment. Changes in relationships or other life events may precipitate depressive episodes; conversely, depressive episodes strain relationships and may lead to negative life events. The goal of treatment is to help the patient solve a crisis in his or her role functioning or social environment. Achieving this helps the patient to gain a sense of mastery over his or her functioning and relieves depressive symptoms.

Begun as a research intervention, IPT has only lately started to be disseminated among clinicians and in residency training programmes. The publication of efficacy data, the promulgation of practice guidelines that embrace IPT among antidepressant treatments, and economic pressures on length of treatment have led to increasing interest in IPT. This chapter describes the concepts and techniques of IPT and its current status of adaptation, efficacy data, and training. The chapter provides a guide to developments and a reference list, but not a comprehensive review.

Background

IPT traces its theoretical and clinical origins to the interpersonal psychoanalytic theory of Meyer and Sullivan and builds on work of other relational theory including object relations, particularly with regard to attachment. However, it applies this theory within a conceptual and clinical framework that differs significantly from that of Sullivan and much of relational theory. In contrast with psychoanalytically inspired schools of thought, IPT sees its goals in treating depression and other medical disorders, rather than trying to change overall personality. Pragmatically, IPT opts to narrow its focus to address the area of interpersonal life that seems to require the most immediate attention.

Acknowledging the importance of personality and early experience, IPT emphasizes the role of recent stressful events in triggering depression and other psychiatric disorders in vulnerable individuals, while it also recognizes the protective role social supports play against psychopathology. Nevertheless, IPT is less interested in discerning the *cause* of a depressive episode (since it assumes the aetiology of depression to be multifactorial) than in using the connection between current life events and the onset of depressive symptoms to help the patient understand and combat the episode of illness.

Compared to other psychotherapies, such as psychoanalytic psychotherapy or even cognitive behavioural therapy, IPT is relatively young. It is less concerned about maintaining an established orthodoxy than about adapting itself to the needs of the patient. Although IPT theorists have taken into account theoretical developments in psychiatry and related fields, much of IPT's evolution has been based on the results of clinical trials. As investigations continue into IPT as a treatment for different disorders and populations, further modification of its theoretical aspects as well as techniques are likely.

Indications

IPT research has demonstrated its efficacy for major depressive disorder across a range of patient ages and contexts, and for bulimia nervosa. One large trial indicates its efficacy (modified as interpersonal social rhythms therapy, or IPSRT) as an adjunctive treatment for bipolar disorder. [6] Lesser evidence suggests the potential benefits of IPT for several anxiety disorders. [21,22] IPT has shown no advantages over control psychotherapies for dysthymic disorder or substance abuse disorders. [14,20] For depressed adolescents, IPT has shown not only efficacy but effectiveness in a school-based programme. [10,11]

Both the physician and patient guides in primary care guidelines for depression list IPT, cognitive behavioural therapy (CBT), behavioural, brief dynamic, and marital therapy as treatments for depression. IPT is spreading from its initial research base in the United States. The IPT manual has been translated into Italian, German, Japanese, Spanish, and French, and is being used ever more widely around the world. Descriptions of IPT have appeared in Spanish, Norwegian, Finnish, and Dutch journals. An International Society for Interpersonal Psychotherapy, established at the American Psychiatric Association Annual meeting in May 2000 in Chicago, has a growing membership and biennial international meetings in 2004 and 2006, and maintains a bibliography of studies.

Because IPT focuses clinically on the social context of the depressive episode, researchers have sometimes adapted IPT when applying it to different treatment populations, developing manuals for different age groups or subpopulations, and occasionally adding focal problem areas. IPT has also been used at different lengths, in different formats, in one pilot couple's adaptation, and as a telephone intervention. Nonetheless, all these adaptations involve the basic principles that constitute IPT: a no-fault definition of the patient's problem as a medical illness, excusing the patient from blame for his or her symptoms; and a continual focus on the relationship between the patient's moods and life situation. The continuing growth of IPT research precludes an exhaustive description of studies. This chapter presents a selection of key research trials of IPT for mood and other disorders (see **Efficacy**) and offers selected references for further reading.

Contraindications

Although formal contraindications (i.e. situations in which IPT would worsen the patients' situation) are not known, IPT was never intended to function as a monotherapy for patients with psychotic depression or bipolar disorder. In addition, three controlled trials have found no benefit of IPT as a treatment for substance use disorders.

Conducting IPT

Each of the four IPT interpersonal problem areas has discrete goals for therapist and patient to pursue. The therapist helps the patient relate life events to mood and other symptoms. In this section we outline the phases of IPT, as well as common strategies and techniques used in IPT treatment. We also outline some differences with cognitive behavioural therapy, to which it is often compared.

Phases of treatment

As an acute treatment, IPT has three phases. The *first phase*, generally covering sessions 1–3, includes diagnostic evaluation,

psychiatric history, and setting the treatment framework. The therapist reviews symptoms, diagnoses the patient as depressed according to DSM-IV (or ICD-10) criteria, and gives the patient the sick role. The psychiatric history includes the 'interpersonal inventory', which is not a structured instrument but a careful review of the patient's past and current social functioning and close relationships, their patterns, and mutual expectations. The relationships are examined to see to what extent they are satisfactory, whether there have been recent changes in those relationships, or whether the patient desires to change them. As part of this review, the therapist commonly links the main social and interpersonal situations of the patient's life to the onset of depressive symptoms.

During the opening phase the therapist also sets a time limit for the acute treatment, generally between 12 and 16 sessions. The optimal number of sessions for IPT requires further research. One study suggests that as few as eight sessions may be effective for some patients, but similar to pharmacological treatment, different doses (i.e. number of IPT sessions) might be necessary for different patients. Sessions are generally scheduled weekly. This allows sufficient time to pass that things will happen in the patient's outside life, on which the treatment focuses. Yet it is frequent enough to maintain momentum and thematic continuity. However, in certain cases logistical difficulties (e.g. due to a general medical illness) might require less frequent sessions.

At the end of the first phase, the therapist links the depressive syndrome to the patient's interpersonal situation focusing on one of the four interpersonal problem areas: (1) *grief*; (2) interpersonal *role disputes*; (3) *role transitions*; or (4) *interpersonal deficits*. Once the patient explicitly accepts this formulation as the focus for treatment, IPT enters its middle phase.

It is important to keep treatment focused on a simple theme. Any formulation necessarily simplifies a patient's life narrative. Although some patients may present with multiple interpersonal problems, the goal of the formulation is to isolate one or, at most, two salient problems related to the patient's mood disorder (whether as precipitant or consequence). More than two foci would risk diffusing the treatment and diluting its efficacy. Sometimes a number of interpersonal problems contribute to the depressive episode, making it apparently difficult to choose a focus. However, research has shown that IPT therapists agree in choosing foci, and patients find those foci credible. Moreover, resolution of the interpersonal treatment focus appears to correlate with symptomatic improvement.

An important task of the initial phase requires deciding whether or not to use medication. A growing literature suggests that combined treatment with antidepressants and IPT works at least as well as, but is not always superior to IPT alone. Thus, except for very severe cases or possibly the elderly, the choice between IPT alone or combined with medication relies more on cost, availability of resources, and patients' preference than on existing empirical evidence.

The *middle phase* involves approaches specific to the chosen interpersonal problem area. For **Grief**—complicated bereavement following the death of a loved one—the therapist facilitates mourning and helps the patient find new activities and relationships to compensate for the loss. **Role disputes** are conflicts with significant others: a spouse, a child, other family members, co-workers, or a close friend. The therapist helps the patient explore the relationship, the nature of the dispute, and available options to negotiate its resolution, including ending the relationship. **Role transition** includes change in life status: for example, beginning or ending a relationship or career, moving, promotion, demotion, retirement, graduation, having a baby, or diagnosis of another medical illness. The patient learns to manage the change by mourning the loss of the old role, recognizing positive and negative aspects of the new role, and taking steps to master this new role.

Interpersonal deficits are used as a focus for patients who lack any of the first three focal life situations. Such patients are isolated or lack social skills, have problems in initiating or sustaining relationships. The goal is to help the patient to develop new relationships and skills. Some patients who fall into this category may in fact suffer from dysthymic disorder or social anxiety disorder, for which separate strategies have been developed.

The *final phase* of IPT, occupying the last 2–4 sessions of acute treatment, builds the patient's newly acquired sense of independence and competence by recognizing and consolidating therapeutic gains. Compared to psychodynamic psychotherapy, IPT de-emphasizes termination: it is a bitter–sweet graduation from successful treatment. The sadness of separating from the therapist is contrasted with depressive feelings.

If the patient has not improved, the therapist emphasizes that the *treatment* has failed, not the patient, and that alternative effective treatments, such as medication or other psychotherapies exist. If the treatment has succeeded, the therapist underscores the patient's competence to function without further therapy by emphasizing that the depressive episode has improved because of the patient's actions in changing a life situation. The therapist also helps the patient to anticipate triggers for and responses to depressive symptoms that might arise in the future.

Patients with multiple prior MDE's or significant residual symptoms, who successfully complete acute treatment but remain at high risk for relapse or recurrence, may contract for maintenance therapy as acute treatment draws to a close. At the end of the treatment (acute or maintenance, depending on the case) the patient is also explicitly told that, should depression recur, the patient should immediately seek treatment, just as the patient would do if any other medical illness recurred.

Techniques

Readers new to IPT will find that much of what we describe below sounds familiar and overlaps with other psychotherapies. Thus, on one level, IPT demands few novel skills from therapists and is relatively easy to learn.

The challenges of IPT lie not in the use of any individual technique, but in organizing these approaches to establish and maintain a coherent primary treatment focus and to resist the temptations of digressing into clinical material outside that focus. Additional challenges may arise from 'unlearning' reflexive responses from prior training experiences such as making transference-focused interventions (for psychodynamic therapists) or identifying automatic cognitions and schemas (for cognitive therapists). In our exposition of strategies and techniques, we focus on major depressive disorder, the first and still best tested indication for IPT, although the same principles may apply to other disorders.

(a) General strategies

IPT is organized around four important concepts.

(i) Psychoeducation

The therapist helps the patient to recognize that the problem is a common medical illness, a mood disorder, with a predictable set of symptoms, not the personal failure or weakness of the patient. IPT therapists define depression as a treatable condition that is not the patient's fault. This definition displaces guilt from the patient to the illness, decreases the patient sense of isolation by feeling part of a larger group (those with depression), and provides hope for a response to treatment.

Underscoring this approach, IPT therapists give depressed patients the 'sick role'. This role temporarily excuses them from what their illness prevents them from doing while assigning them the task of working as patients in order to recover their previous healthy role. The resolution of the sick role is to regain the healthy, euthymic role by the end of treatment. The time-limited structure of IPT also energizes patients and protects against regression during treatment.

(ii) Focusing on the positive

IPT therapists take an empathic, supportive, and encouraging stance. They emphasize their patients' successes, although they also commiserate on their difficulties. 'Focusing on the positive' means underscoring positive events; it does *not* mean ignoring negative affect. By doing this, IPT therapists may facilitate the therapeutic alliance that is crucial to good outcome. By solving an interpersonal crisis—a complicated bereavement, a role dispute or transition, or an interpersonal deficit—the IPT patient has the dual opportunity to improve his or her life situation and simultaneously relieve the symptoms of the depressive episode.

This coupled formula, validated by randomized controlled trials in which IPT has been tested, can be offered with confidence and optimism. Symptomatic relief may correlate with the degree to which the patient solves his or her interpersonal crisis. This therapeutic optimism, while not specific to IPT, very likely provides part of its power in remoralizing the patient.

(iii) Focus on the present, not the past

IPT deals with current rather than past interpersonal relationships, focusing on the patient's immediate social context. The IPT therapist attempts to intervene in depressive symptom formation and social dysfunction rather than addressing enduring aspects of personality, which are difficult to assess accurately during an episode of an Axis I disorder. However, IPT does build new social skills such as self-assertion and increased ability to understand interpersonal exchanges, which may be as valuable as changing personality traits.

(iv) Link mood to life events

A core strategy of IPT is constant attention to the link between the patient's current mood state and recent interpersonal experiences. Stressful life events and negative interpersonal encounters trigger lower mood and can lead to depressive episodes in vulnerable individuals. Conversely, depressed mood impairs social functioning, which can lead to further negative life events. IPT is postulated to work by helping the patients manage interpersonal relationships more effectively, which leads to improved mood. Improved mood then allows patients to more effectively manage interpersonal experiences in an iterative fashion.

Unlike psychodynamic psychotherapy, IPT *does not* focus on early childhood experiences and long-standing familial dynamics. Thus, the patient's current mood state is linked to recent experiences rather than those rooted in the distant past. Nor does IPT focus on transferential material, except in the relatively rare instance when problems arise in the therapeutic alliance. Thus the treatment highlights recent experiences outside the office.

(b) Specific techniques

To achieve the general goals of IPT, the following techniques are frequently used:

1 *An opening question*: 'how have things been since we last met?', which leads the patient to provide an interval history of mood and events. The therapist begins each session after the first one with this tactic. It is common, particularly at the beginning of the treatment that the patient will focus exclusively either on the mood or on a recent event. When that occurs, the therapist gently asks about the other aspect and helps the patient connect mood and recent events.

2 *Communication analysis*, a detailed recreation of recent, affectively charged circumstances. This detailed analysis often helps the patient uncover nuances of the interpersonal exchanges that had been missed prior to the session.

3 An exploration of the patient's wishes and options, to help the patient realize and voice the desired outcomes.

4 *Decision analysis*, to help the patient integrate communication analysis, the wishes and options and the constraints of the situation and decide on a specific course of action.

5 *Role-playing*, to help the patient rehearse that course of action before implementing in real life.

Similarities and differences with other psychotherapies

Because IPT and CBT are the two best empirically supported psychotherapies, they are often compared. [23] IPT shares with CBT an orientation towards making the patient feel understood, a 'here and now' focus, a general feeling of hope and optimism, psychoeducation, and the use of role-playing to favour the acquisition of new skills. It addresses interpersonal issues in a manner familiar to marital therapists. Although like CBT a time-limited treatment targeting a syndromal constellation (e.g. major depression), IPT is considerably less structured, and focuses on interpersonal problem areas rather than automatic thoughts. IPT overlaps to some degree with psychodynamic psychotherapies, yet IPT also meaningfully differs from them: in its focus on the present, not the past; its focus on practical, real-life change rather than self-understanding; its medical model; and its avoidance of the transference and of genetic and dream interpretations.

Efficacy

Research findings: IPT for mood disorders

IPT outcome research is ongoing, with new studies published every year. What follows is a selection of key research trials of IPT for mood and other disorders. For some of these trials, IPT was adapted in a separate treatment manual, but in all cases the general principles of the treatment remained the same.

(a) Acute treatment of major depression

IPT was first tested as an acute antidepressant treatment in a four-cell, 16-week randomized trial. This compared weekly IPT, amitriptyline (AMI), their combination, and a monthly supportive psychotherapy treatment for 81 outpatients with major depression.[24,25] The outcome of patients receiving amitriptyline and IPT was similar and superior to that of supportive psychotherapy. Patients who received both amitriptyline and IPT had better depression outcomes and better scores on a range of social adjustment measures including overall adjustment, work performance, and communication than those on amitriptyline alone, suggesting an additive effect of IPT on medication treatment. At 1-year follow-up, many patients sustained improvement from the brief IPT intervention, and IPT patients had developed significantly better psychosocial functioning whether or not they received medication. This effect on social function was not found for AMI alone and had not been evident for IPT at the end of the 16-week trial.

Still probably the most important study to date involving IPT is the National Institute of Mental Health Treatment of Depression[24,25] Collaborative Research Program (TDCRP), investigators randomly assigned 250 outpatients with major depression to 16 weeks of IPT, CBT, or either imipramine (IMI) or placebo with clinical management.[2] Most subjects completed at least 15 weeks or 12 treatment sessions. Patients with milder depression (defined as a 17-item Hamilton Depression Rating Scale [HDRS] score <20) improved equally in all four treatments. For more severely depressed patients (HDRS>20), IMI worked fastest and most consistently better than placebo. IPT and IMI were comparable on several outcome measures, including HDRS, and superior to placebo for more severely depressed patients. In some analyses, IPT appeared to be slightly superior to CBT. CBT was not superior to placebo among the more depressed patients.

A follow-up study of TDCRP subjects 18 months later found no significant difference in recovery among remitters (who had minimal or no symptoms after the end of treatment, sustained during follow-up) among the four treatments.[26] Thirty per cent of CBT, 26 per cent of IPT, 19 per cent of imipramine, and 20 per cent of placebo subjects initially randomized to those treatments remitted and remained in remission during that time span. Among acute remitters, relapse over the 18-month follow-up was 36 per cent for CBT, 33 per cent for IPT, 50 per cent for imipramine (medication having been stopped at 16 weeks), and 33 per cent for placebo. The authors concluded that, for many patients, 16 weeks of specific treatments were insufficient to achieve full and lasting recovery.

Special populations and settings

(a) Depressed primary care patients

There has been a study comparing IPT and nortriptyline with usual care for depressed patients in a primary care setting. If patients were hospitalized for a general medical condition, IPT was continued in the hospital when possible. Depressive symptom severity declined more rapidly with either nortriptyline or IPT than in usual care. Approximately 70 per cent of treatment completers receiving nortriptyline or IPT, but only 20 per cent in usual care, had recovered after 8 months. Subjects with a lifetime history of comorbid panic disorder had a poorer response across treatments, compared to those with major depression alone.

(b) Depressed HIV-positive patients (IPT-HIV)

IPT has also been investigated for depressed HIV patients (IPT-HIV), emphasizing common issues among this population including concerns about illness and death, grief and role transitions. A randomized study echoing the TDCRP of 101 HIV-positive patients with depressive symptoms found IPT and imipramine each superior to CBT and supportive therapy. Many patients reported improvement in depressive physical symptoms that they had mistakenly attributed to HIV infection. IPT may have been a better fit than CBT for these patients due to the extreme life events they faced at the height of the HIV epidemic.

(c) Peripartum depression

Pregnancy and the postpartum also provide natural role transitions as an IPT focus. Exploring these role transitions addresses the depressed pregnant woman's self-evaluation as a parent, physiologic changes of pregnancy and postpartum, and altered relationships with the spouse or significant other and with other children. Timing and duration of sessions are adjusted in response to bedrest, delivery, obstetrical complications, and childcare. Postpartum mothers may bring children to sessions. As with depressed HIV-positive patients, telephone sessions, and hospital visits are sometimes necessary. A controlled clinical trial comparing IPT to a didactic parent education group in depressed pregnant women showed advantages for IPT. A study of depressed postpartum women found superiority of IPT over a wait-list control group. A small randomized trial has also suggested the possibility that group IPT may serve to prevent MDD relapse during the postpartum period.

(d) Conjoint IPT for depressed patients with marital disputes (IPT-CM)

Marital conflict can precipitate or complicate depressive episodes. Some clinicians believe that individual psychotherapy for patients in marital disputes may lead to premature rupture of marriages. Researchers at Yale University developed a manual for conjoint therapy of depressed patients with marital disputes (IPT-CM). IPT-CM includes the spouse in all sessions and focuses on the current marital dispute. Eighteen patients with major depression linked to the onset or exacerbation of marital disputes were randomly assigned to 16 weeks of either individual IPT or IPT-CM. Patients in both treatments showed similar reductions in depressive symptoms, but patients receiving IPT-CM reported significantly better marital adjustment, marital affection, and sexual relations. These pilot findings require replication with a larger sample and other control groups.

(e) Depressed adolescents (IPT-A)

IPT has also been modified to incorporate adolescent developmental issues. Three randomized trials, one of them conducted in Puerto Rico, have shown the efficacy of IPT-A. It is important to note that in the Puerto Rico study, the only one that also included CBT in the design, IPT appeared superior to CBT in certain measures (e.g. self-esteem and social adaptation), consistent with the findings of the TDCRP and the study of depressed HIV-positive patients.

(f) Maintenance treatment

Based on the success of IPT as acute treatment for MDD, the recurrent nature of mood disorders, and the efficacy of medication in preventing relapse and recurrence, IPT was adapted as a once

monthly maintenance treatment for MDD (IPT-M). This was a novel development and allowed the first real testing of psychotherapy as a maintenance treatment for patients who had remitted from acute depression. Since IPT-M begins with patients who have remitted, its goal is to maintain the remitted state. Both patient and therapist are vigilant for early signs of interpersonal problems similar to those of which the patient and therapist previously identified as associated with the onset of the patient's most recent depressive episode. At the same time, the therapist works to enhance strengths that appear to have been present prior to the patient's illness or began to emerge as the most recent depressive episode remitted. In contrast with the acute phase application of IPT, which usually focuses on one or at most two interpersonal problem areas, IPT-M may shift problem areas over time. Three studies have compared medication and IPT as maintenance treatment for MDD.

In the first study, 128 outpatients with recurrent depression initially treated with combined high dose (>200 mg/day) imipramine and weekly sessions of IPT. [4,5] Responders remained on high-dosage medication while IPT was tapered to a monthly frequency during a 4-month continuation phase. Patients who remained remitted were then randomly assigned to 3 years of either: (1) imipramine plus clinical management; (2) imipramine plus monthly IPT; (3) monthly IPT alone; (4) monthly IPT plus placebo; or (5) placebo plus clinical management.

Both IPT and imipramine were significantly superior to the placebo group in delaying MDD relapse. Imipramine was superior to IPT-M in the ability to prevent relapse. The group that received IPT with imipramine had a numerically lower rate of recurrence at 1 year (16 per cent) than the group on imipramine alone (40 per cent), but those results were not statistically significant. Two different studies have had very similar findings in comparisons of IPT and nortriptyline for geriatric patients with recurrent major depression. [7, 27]

Further study is required to determine the efficacy of IPT relative to newer medications (e.g. selective serotonin-reuptake inhibitors), and the efficacy of dosages other than once monthly maintenance IPT. A study of differing doses of maintenance IPT for depressed patients in Pittsburgh has not found differences in outcome based on frequency of sessions.[28] Perhaps optimal dosing of maintenance IPT depends on individual patients' needs.

The success of IPT in treating MDD has led researchers to investigate its efficacy in bipolar and dysthymic disorder.

(i) Bipolar disorder

The modification of IPT used as an adjunct to medication in the treatment of bipolar disorder is called interpersonal and social rhythm therapy (IPSRT). Its use rests on the hypothesis that disruptions of social rhythms are destabilizing for bipolar patients and contribute to trigger their relapse. By decreasing the number and intensity of those disruptions, IPSRT should improve the course of bipolar disorder. The behavioural component helps to protect sleep patterns and limit the disruptions that may provoke mania; the IPT approach to depression remains largely the same.

After stabilizing bipolar I patients with appropriate pharmacotherapy and either IPSRT or intensive clinical management, patients were randomized again to either IPSRT or clinical management for preventive treatment.[6] They found that participants assigned to IPSRT acutely had longer survival times to a new affective episode, irrespective of maintenance treatment assignment. Participants in the IPSRT group had higher regularity of social rhythms at the end of the acute treatment, and this increased regularity of social rhythms during the acute treatment mediated the reduced likelihood of recurrence during the maintenance treatment. Further research appears necessary to more firmly establish the optimal timing and treatment duration of IPSRT.

(ii) Dysthymic disorder

A modification of IPT for dysthymic disorder[29] encourages patients to reconceptualize what they have considered their lifelong character flaws as ego-dystonic, chronic mood-dependent symptoms: as chronic but treatable 'state' rather than immutable 'trait'.

Three randomized trials have examined the efficacy of IPT in dysthymic disorder. In the first study, 35 patients with an ICD-10 diagnosis of dysthymia with or without comorbid MDD were randomized to moclobemide alone ($n = 19$) or moclobemide plus IPT ($n = 16$). Patients were assessed with the 17-item Hamilton Rating Scales for Depression (HAM-D), the Global Assessment Scale (GAS), and the Quality of Life and Satisfaction Questionnaire at baseline, 12, 24, and 48 weeks. Both groups showed statistically significant improvement in all measures across time. There were no differences between the two treatments at week 12. However, patients in the combined group had statistically better scores than the patients in the moclobemide group on all outcome variables at weeks 24 and 48.

In the second study, 707 adults in primary care clinic with DSM-IV dysthymic disorder were randomized, with or without past and/or current MDD (15 per cent of the sample had current MDD), to treatment with sertraline alone (50–200 mg), IPT alone (10 sessions), or sertraline with IPT combined.[30] At the end of treatment, response rates were 60 per cent for sertraline alone, 47 per cent for IPT alone, and 58 per cent for sertraline with IPT. After an additional 18-month naturalistic follow-up phase, there were no statistically significant differences in symptom reduction between sertraline alone and sertraline with IPT. However, both were more effective than IPT alone in reducing depressive symptoms. It is important to note, though, that IPT was given as a brief treatment, while sertraline was generally continued for the full 2 years of the study.

A third study compared IPT adapted for the treatment of dysthymia (IPT-D), brief supportive psychotherapy, sertraline, and combined IPT-D/sertraline for patients with pure dysthymic disorder (i.e. without 'double' depression) in 94 subjects treated over 16 weeks.[14] Patients improved in all conditions, with the cells including sertraline pharmacotherapy showing superiority over psychotherapy alone in response and remission. The results of this study are consistent with an emerging literature suggesting that pharmacotherapy may acutely benefit patients more than psychotherapy. In conjunction with the other studies in patients with dysthymic disorder, it suggests that IPT alone may not be an efficacious treatment for dysthymic disorder.

IPT for non-mood disorders

The efficacy of IPT as an antidepressant treatment has led to its adaptation as a treatment for other psychiatric disorders, based on the premise that life events are ubiquitous.

(a) Bulimia

Fairburn *et al.* modified IPT for the treatment of bulimia, eliminating the use of the sick role and of role-playing in order to contrast

distinct therapeutic strategies in comparing IPT and CBT. Initial trials showed that although CBT worked faster to relieve bulimic symptoms, IPT had longer-term benefits comparable to CBT and superior to a behavioural control condition.[1] A subsequent multisite trial found CBT superior to IPT.[9]

Following a model closer to the original IPT principles, Wilfley *et al*. modified IPT in a group format (IPT-G) and compared it to group CBT and a wait-list control for 56 women with non-purging bulimia.[8] The initial IPT phase was conducted individually. The interpersonal area for almost of all subjects was formulated as 'interpersonal deficits'. At termination, binge eating decreased in the IPT-G and CBT groups, but not in the control condition. Results persisted at 1-year follow-up. A randomized clinical trial of 162 women, comparing group IPT, and CBT for 20 sessions over 20 weeks, yielded similar results.

A research group in Christchurch, New Zealand studied the application of IPT to anorexia nervosa.[19] In their trial, neither IPT nor CBT showed efficacy as an outpatient treatment, consistent with the general anorexia outcome literature.

(b) Anxiety disorders

IPT has not yet been tested in controlled studies for anxiety disorders. Promising results have been found in for social anxiety disorders, PTSD, and panic disorder.[22, 31–33] Several groups are currently conducting controlled trials for these disorders.

(c) Substance abuse

IPT has failed to demonstrate efficacy in three clinical trials for patients with substance dependence.[20, 34, 35] These negative studies suggest limits to the range of utility of IPT as a main treatment for substance use disorders, but do not necessarily preclude its use to treat MDD comorbidity in those patients.

(d) Other applications

Research groups are testing the applicability of IPT to body dysmorphic disorder, chronic somatization in primary care patients, depressed patients postmyocardial infarction, depressed cancer patients, borderline personality disorder, insomnia, and other disorders. The IPT focus on life events suggests its potential applicability to patients with medical illness.

(e) IPT by telephone

Because many patients avoid or have difficulty reaching an office for face-to-face treatment, IPT and IPC are being tested as a treatment delivered over the telephone. Weissman and Miller conducted a successful pilot feasibility trial comparing IPT by telephone to wait-list control in 30 patients with recurrent major depression, and found IPT to be the superior control condition in reducing depressive symptoms and improving psychosocial functioning.[36] Neugebauer and colleagues found telephone IPC a helpful intervention for women with subsyndromal depression following a miscarriage.[37, 38]

(f) Interpersonal counselling (IPC)

Many patients presenting for treatment, particularly outside mental health settings, report psychiatric symptoms but do not meet threshold criteria for a psychiatric disorder. Nonetheless, their symptoms can be debilitating, interfering with their daily functioning, and often result in increased use of general medical services. Interpersonal counselling (IPC), based on IPT, was designed to treat distressed primary care patients who do not meet full syndromal criteria for psychiatric disorders. IPC is administered for a maximum of six sessions by health care usually by professionals who lack formal psychiatric training such as nurse practitioners. The first session can last up to 30 min; subsequent sessions are briefer.

IPC therapists assess the patient's current functioning, recent life events, occupational and familial stressors, and changes in interpersonal relationships. They assume that such events provide the context in which emotional and bodily symptoms occur. Klerman and colleagues studied 128 patients in a primary care clinic who scored 6 or higher on the Goldberg General Health Questionnaire (GHQ), randomizing them to IPC or to usual care without psychological treatment.[39] Over an average of 3 months, often receiving only one or two IPC sessions, IPC subjects showed significantly greater symptom relief on the GHQ than controls, especially mood improvement. IPC subjects were more likely to subsequently make use of mental health services, suggesting a new awareness of the psychological aspect of their symptoms.

Predictors of response to IPT

Five studies have examined predictors of response to IPT. Analyses of the TDCRP data identified general predictors of response to MDD treatment, as well as predictors for specific treatment modalities.[40,41] Seven patient characteristics predicted outcome across treatments: social dysfunction (higher social function predicted better response), cognitive dysfunction (better cognitive function predicted better response, particularly to CBT), expectation of improvement (higher expectation predicted better response), therapeutic alliance (a stronger alliance predicted better response), endogeneity of depression (endogenous depression tended to have better response, a finding supported by another study examining the relationship between EEG patterns and response to IPT in depressed patients[42]), double depression (its presence predicted poorer outcome), personality traits (their presence predicted worse response), and duration of current episode (longer duration was associated to worse response). In addition prior social adjustment, as measured by previous attainment of a marital relationship and higher satisfaction with social relationships in general, differentially predicted good response to IPT. This finding is consistent with reports in the general psychotherapy literature documenting that various indicators higher baseline psychosocial functioning predict good psychotherapy response. Two other studies suggest that comorbidity tends to worsen the prognosis of treatment with IPT alone, but not the prognosis of combined treatment.[43,44]

Summary of research findings

IPT has demonstrated efficacy as an acute and maintenance monotherapy and as a component of combined treatment for major depressive disorder. It also appears to have utility for other mood and non-mood syndromes, although the evidence for these is sparser. It has not shown benefit for substance use disorders or as a monotherapy for dysthymic disorder. Since monotherapy with either IPT or pharmacotherapy is likely to suffice for most patients with major depressive disorder, combined treatment is probably best reserved for severely or chronically ill patients. How best to combine time-limited psychotherapy with pharmacotherapy—for which patients, in what sequence, etc.—is an exciting area for future research.

Training

Until recently, IPT therapists were few, and practiced almost exclusively in research studies. Publications supporting its efficacy have led to clinical demand for this empirically supported treatment. IPT training is now increasingly included in professional workshops and conferences, with training courses conducted at University centres in Canada, the United Kingdom, continental Europe, Asia, New Zealand, and Australia. IPT is taught in a small but growing minority of psychiatric residency training programmes in the United States and as well as some family practice and primary care training programmes.

Although the principles and practice of IPT are relatively straightforward, any psychotherapy requires innate therapeutic ability, comfort with the so-called common factors of psychotherapy: tolerating and exploring affect, helping the patient to feel understood, engendering hope, etc. IPT training requires more than reading the manual: psychotherapy is learned by doing. Most IPT training programmes are designed to help experienced therapists refocus their treatment by learning new techniques, not to teach novices psychotherapy. Candidates should have a graduate clinical degree (MD, Ph.D., MSW, RN), several years of experience conducting psychotherapy, and clinical familiarity with the diagnosis of patients they plan to treat.

The IPT training in the TDCRP became the model for subsequent research studies. It included a brief didactic programme, reading the manual, and a practicum in which the therapist treated 2–3 patients under close supervision monitored by videotapes of the sessions. For research certification, we continue to recommend at least two or three successfully treated cases with hour for hour supervision of taped sessions.

Although many clinicians would like a formal certificate or diploma in IPT, there is no gold standard of IPT proficiency and no accrediting board. When IPT practice was limited to research settings, this posed no problem: one research group taught another, in the manner described above. As IPT spreads in clinical practice, the educational and accreditation process for IPT requires further study. The newly created International Society for Interpersonal Psychotherapy (ISIPT) may provide an appropriate forum in which to discuss these increasingly important issues.

Future directions

The history of IPT has been a succession of outcome trials. These studies have helped to define diagnostic indications for this treatment, but we know far less about the dosage and indications of IPT than about antidepressant medication. Future outcome trials may continue to define the scope of efficacy (response to treatment under ideal conditions) and effectiveness (response to treatment in more general clinical settings) of IPT. These should include both tests for different diagnoses, such as the anxiety disorders, testing of dosage—optimal frequency and duration of IPT sessions—and also studies of the sequencing of IPT with other treatments. Other research may help to determine the cost-effectiveness and potential cost-offset of IPT as a treatment that improves both symptoms and social functioning.

Most of the work on IPT to date has focused on treatment outcome. By contrast, little is known about the process aspects in IPT such as the specific value of many IPT interventions. Although it appears that solving an interpersonal problem area correlates with treatment outcome, it is unclear, for example, whether the choice of a particular treatment focus over other makes a difference for patients, or whether particular sorts of life events are helpful or unhelpful foci. Patient and therapist characteristics may also potentially influence treatment outcome.

Finally, while the initial work on IPT was conducted in the United States, over the last few years, IPT trials have also been conducted in other countries. As those studies continue to be conducted, it will become easier to discern to what extent IPT addresses topics are universal across cultures.

In summary, IPT is a time-limited, forward-looking, pragmatically focused psychotherapy that defines psychiatric disorders as treatable medical illnesses and links them to the patient's current social situation. This strategy has proved efficacious for patients with major depression and bulimia, and shows promise for other mood and non-mood disorders.

Further information

Elkin, I., Shea, M.T., Watkins, J.T., *et al.* (1989). National Institute of Mental Health treatment of depression collaborative research program: general effectiveness of treatments. *Archives of General Psychiatry*, **46**, 971–82.

Weissman, M.M., Markowitz, J.C., and Klerman, G.L. (2000). *Comprehensive guide to interpersonal psychotherapy*. Basic Books, New York.

Weissman, M.M., Markowitz, J.C., and Klerman, G.L. (2007). *Clinician's quick guide to interpersonal psychotherapy*. Oxford University Press, New York.

Website of the International Society of Interpersonal Psychotherapy: http://www.interpersonalpsychotherapy.org/

References

1. Fairburn, C.G., Jones, R., Peveler, R.C., *et al.* (1993). Psychotherapy and bulimia nervosa. Longer-term effects of interpersonal psychotherapy, behavior therapy, and cognitive behavior therapy. *Archives of General Psychiatry*, **50**(6), 419–28.
2. Elkin, I., Shea, M.T., Watkins, J.T., *et al.* (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Archives of General Psychiatry*, **46**(11), 971–82; discussion 983.
3. Klerman, G.L. and Weissmann, M.M. (1987). Interpersonal psychotherapy (IPT) and drugs in the treatment of depression. *Pharmacopsychiatry*, **20**(1), 3–7.
4. Frank, E., Kupfer, D.J., Wagner, E.F., *et al.* (1991). Efficacy of interpersonal psychotherapy as a maintenance treatment for recurrent depression: contributing factors. *Archives of General Psychiatry*, **48**, 1053–9.
5. Frank, E., Kupfer, D.J., Perel, J.M., *et al.* (1990). Three-year outcomes for maintenance therapies in recurrent depression. *Archives of General Psychiatry*, **47**(12), 1093–9.
6. Frank, E., Kupfer, D.J., Thase, M.E., *et al.* (2005). Two-year outcomes for interpersonal and social rhythm therapy in individuals with bipolar I disorder. *Archives of General Psychiatry*, **62**(9), 996–1004.
7. Reynolds, C.F. III, Frank, E., Perel, J.M., *et al.* (1999). Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression: a randomized controlled trial in patients older than 59 years. *The Journal of the American Medical Association*, **281**(1), 39–45.
8. Wilfley, D.E., Welch, R.R., Stein, R.L., *et al.* (2002). A randomized comparison of group cognitive-behavioral therapy and group interpersonal psychotherapy for the treatment of overweight individuals with binge-eating disorder. *Archives of General Psychiatry*, **59**(8), 713–21.

9. Agras, W.S., Walsh, B.T., Fairburn, C.G., *et al.* (2000). A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *Archives of General Psychiatry,* **57**(5), 459–66.
10. Mufson, L., Dorta, K.P., Wickramaratne, P., *et al.* (2004). A randomized effectiveness trial of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry*, **61**, 577–84.
11. Mufson, L., Weissman, M.M., Moreau, D., *et al.* (1999). Efficacy of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry*, **56**(6), 573–9.
12. Frank, E., Cyranowski, J.M., Rucci, P., *et al.* (2002). Clinical significance of lifetime panic spectrum symptoms in the treatment of patients with bipolar I disorder. *Archives of General Psychiatry*, **59**(10), 905–11.
13. Markowitz, J.C. (1994). Psychotherapy of dysthymia. *American Journal of Psychiatry*, **151**, 1114–21.
14. Markowitz, J.C., Kocsis, J.H., Bleiberg, K.L., *et al.* (2005). A comparative trial of psychotherapy and pharmacotherapy for "pure" dysthymic patients. *Journal of Affective Disorders*, **89**(1–3), 167–75.
15. Markowitz, J.C., Svartberg, M., and Swartz, H.A. (1998). Treatment of HIV-positive patients with depressive symptoms. *Archives of General Psychiatry*, **55**, 452–7.
16. Spinelli, M.G. (1997). Interpersonal psychotherapy for depressed antepartum women: a pilot study. *The American Journal of Psychiatry*, **154**(7), 1028–30.
17. Spinelli, M.G. and Endicott, J. (2003). Controlled clinical trial of interpersonal psychotherapy versus parenting education program for depressed pregnant women. *The American Journal of Psychiatry*, **160**(3), 555–62.
18. Schulberg, H.C., Block, M.R., Madonia, M.J., *et al.* (1996). Treating major depression in primary care practice. Eight-month clinical outcomes. *Archives of General Psychiatry*, **53**(10), 913–9.
19. McIntosh, V.V., Jordan, J., Carter, F.A., *et al.* (2005). Three psychotherapies for anorexia nervosa: a randomized, controlled trial. *The American Journal of Psychiatry*, **162**(4), 741–7.
20. Rounsaville, B.J., Glazer, W., Wilber, C.H., *et al.* (1983). Short-term interpersonal psychotherapy in methadone-maintained opiate addicts. *Archives of General Psychiatry*, **40**(6), 629–36.
21. Lipsitz, J.D., Fyer, A.J., Markowitz, J.C., *et al.* (1999). Open trial of interpersonal psychotherapy for the treatment of social phobia. *The American Journal of Psychiatry*, **156**(11), 1814–16.
22. Bleiberg, K.L. and Markowitz, J.C. (2005). A pilot study of interpersonal psychotherapy for posttraumatic stress disorder. *The American Journal of Psychiatry*, **162**(1), 181–3.
23. Markowitz, J.C., Svartberg, M., and Swartz, H.A. (1998). Is IPT time-limited psychodynamic psychotherapy? *Journal of Psychother Practice and Research*, **7**(3), 185–95.
24. DiMascio, A., Weissman, M.M., Prusoff, B.A., *et al.* (1979). Differential symptom reduction by drugs and psychotherapy in acute depression. *Archives of General Psychiatry*, **36**(13), 1450–6.
25. Weissman, M.M. (1979). The psychological treatment of depression. Evidence for the efficacy of psychotherapy alone, in comparison with, and in combination with pharmacotherapy. *Archives of General Psychiatry*, **36**(11), 1261–9.
26. Shea, M.T., Elkin, I., Imber, S.D., *et al.* (1992). Course of depressive symptoms over follow-up. Findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry*, **49**(10), 782–7.
27. Reynolds, C.F. III, Frank. E., Kupfer, D.J., *et al.* (1996). Treatment outcome in recurrent major depression: a post hoc comparison of elderly ("young old") and midlife patients. *The American Journal of Psychiatry*, **153**(10), 1288–92.
28. Frank, E., Kupfer, D.J., Buysse, D.J., *et al.* (2007). Randomized trial of weekly, twice-monthly, and monthly interpersonal psychotherapy as maintenance treatment for women with recurrent depression. *The American Journal of Psychiatry*, **164**(5), 761–7.
29. De Mello, M.F., Myczcowisk, L.M., and Menezes, P.R. (2001). A randomized controlled trial comparing moclobemide and moclobemide plus interpersonal psychotherapy in the treatment of dysthymic disorder. *Journal of Psychother Practice Research*, **10**(2), 117–23.
30. Browne, G., Steiner, M., Roberts, J., *et al.* (2002). Sertraline and/or interpersonal psychotherapy for patients with dysthymic disorder in primary care: 6-month comparison with longitudinal 2-year follow-up of effectiveness and costs. *Journal of Affective Disorders*, **68**(2–3), 317–30.
31. Lipsitz, J.D., Gur, M., Miller, N.L., *et al.* (2006). An open pilot study of interpersonal psychotherapy for panic disorder (IPT-PD). *The Journal of Nervous and Mental Disease*, **194**(6), 440–5.
32. Lipsitz, J.D., Gur, M., Vermes, D., *et al.* (2007). A randomized trial of interpersonal therapy versus supportive therapy for social anxiety disorder. *Depression and Anxiety*, **25**(6), 542–53
33. Lipsitz, J.D., Fyer, A.J., Markowitz, J.C, (1999). Open trial of interpersonal psychotherapy for the treatment of social phobia. *The American Journal of Psychiatry*, **156**(11), 1814–6.
34. Carroll, K.M., Fenton, L.R., Ball, S.A., *et al.* (2004). Efficacy of disulfiram and cognitive behavior therapy in cocaine-dependent outpatients: a randomized placebo-controlled trial. *Archives of General Psychiatry*, **61**(3), 264–72.
35. Carroll, K.M., Rounsaville, B.J., and Gawin, F.H. (1991). A comparative trial of psychotherapies for ambulatory cocaine abusers: relapse prevention and interpersonal psychotherapy. *The American Journal of Drug Alcohol Abuse*, **17**(3), 229–47.
36. Miller, L. and Weissman, M. (2002). Interpersonal psychotherapy delivered over the telephone to recurrent depressives. A pilot study. *Depression and Anxiety*, **16**(3), 114–7.
37. Neugebauer, R., Kilne, J., Bleiberg, K., *et al.* (2007). Preliminary open trial of interpersonal counseling for subsyndromal depression following miscarriage. *Depression and Anxiety*, **24**(3), 219–22.
38. Neugebauer, R., Kline, J., Markowitz, J.C., *et al.* (2006). Pilot randomized controlled trial of interpersonal counseling for subsyndromal depression following miscarriage. *The Journal of Clinical Psychiatry*, **67**(8), 1299–304.
39. Klerman, G.L., Budman, S., Berwisk, D., *et al.* (1987). Efficacy of a brief psychosocial intervention for symptoms of stress and distress among patients in primary care. *Medical Care*, **25**(11), 1078–88.
40. Sotsky, S.M., Glasss, D.J., Shear, M.T., *et al.* (1991). Patient predictors of response to psychotherapy and pharmacotherapy: findings in the NIMH Treatment of Depression Collaborative Research Program. *The American Journal of Psychiatry*, **148**(8), 997–1008.
41. Barber, J.P. and Muenz, L.R. (1996). The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: empirical findings from the treatment for depression collaborative research program. *Journal of Consulting & Clinical Psychology*, **64**(5), 951–8.
42. Thase, M.E., Buysse, D.J., Frank, E., *et al.* (1997). Which depressed patients will respond to interpersonal psychotherapy? The role of abnormal EEG sleep profiles. *The American Journal of Psychiatry*, **154**(4), 502–9.
43. Frank, E., Shear, M.K., Rucci, P., *et al.* (2000). Influence of panic-agoraphobic spectrum symptoms on treatment response in patients with recurrent major depression. *The American Journal of Psychiatry*, **157**(7), 1101–7.
44. Brown, C., Suhulberg, H.C., Madonia, M.J., *et al.* (1996). Treatment outcomes for primary care patients with major depression and lifetime anxiety disorders. *The American Journal of Psychiatry*, **153**(10), 1293–300.

6.3.4 Brief individual psychodynamic psychotherapy

Amy M. Ursano and Robert J. Ursano

Introduction

Interest in brief dynamic psychotherapy has flourished in recent years. The psychodynamic psychotherapies, including brief psychodynamic psychotherapy, aim to change behaviour through new understanding and the recognition of maladaptive patterns of behaviour enacted since childhood but not previously observed. Through this process, perceptions, expectations, beliefs, and, therefore, behaviours and feelings are altered.[1]

Historically, 'brief psychotherapy' and 'long-term psychotherapy' were used synonymously with 'supportive' and 'explorative' psychotherapy, respectively. However, brief and long-term describe only the duration rather than the technique, focus, or goal of treatment.[2] The time limits of brief dynamic psychotherapy give it a unique character and distinguish it from long-term psychotherapy and psychoanalysis. Because of its limited goals, the brief dynamic psychotherapist must confront his or her ambitiousness and perfectionism as well as any exaggerated ideal of personality structure and function.

Psychotherapy in general, and brief individual psychodynamic psychotherapy in particular, is perhaps the most elegant form of micro-neurosurgery. Psychotherapy strives to alter behaviour (i.e. cognitions, affects, and actions) with verbal interchange—fundamentally to change neurone A that used to connect to neurone B so it will now connect to neurone C. Although the therapist in the individual psychodynamically derived psychotherapies does not 'require' behavioural change, the end result of the therapist's technical expertise is to achieve behavioural change, including changes in well-being, physical health, social supports, and societal productivity as well as symptomatic relief. As in all of medicine, both non-specific and specific curative factors affect the outcome of this work. The non-specific curative factors—abreaction, the provision of new information, and maximizing success experiences—are present in all forms of medical treatment including brief psychotherapy. Brief individual dynamic psychotherapy also has specific technical interventions and procedures above and beyond the non-specific curative factors. As in other medical therapies, there are contraindications and dangers in the use of this treatment.

Background

Evolving from psychoanalysis in the mid-twentieth century, brief individual psychodynamic psychotherapy, like other psychodynamic treatments, is based on the principle that meanings and past experience play an important role in behaviour and illness. Although psychoanalysis is now a lengthy procedure usually requiring a number of years to complete, the early psychoanalytic literature, including Freud's first cases, contain histories of successful short analyses. During the first 30 years of psychoanalysis, it was unusual for treatments to extend beyond 1 year.[3] Ferenczi was the first analyst to advocate shortening psychoanalysis. He advocated 'active therapy' a more directive, focused, and briefer treatment. Rank was the first one to explicitly to set a time limit on treatment. Ferenczi and Rank[4] articulated the advantages of brief dynamic psychotherapy.

Following the Second World War, the interest in psychoanalysis resulted in greater demand for psychotherapy and increased pressure to develop briefer treatments. In the mid-1940s, Alexander and French advocated shortening treatment by decreasing the frequency of sessions in order to minimize regression. They proposed to focus treatment on the present rather than the past, using historical conflicts to inform the therapist in providing the best corrective emotional experience for the patient in the present.

The community-based mental health treatment movement, the increasing cost of mental health care, and the rise of managed care in the United States; have stimulated efforts to find briefer forms of psychotherapy. Contemporary brief individual psychodynamic psychotherapy is heavily influenced by the British School's development of brief focal psychotherapy. Balint sponsored a workshop of experienced psychoanalytic psychotherapists, which focused on clinical evaluation and attempted to understand which patients might be suitable for briefer treatment. After Balint's death, Malan carried on the work of the group. At the Tavistock clinic, Malan developed and applied the principles of psychodynamic treatment to brief treatment, delineating methods for evaluating process and outcome variables. He emphasized the importance of therapeutic planning and the identification of a focal conflict.

Concurrently, Sifneos, at the Massachusetts General Hospital, was studying brief psychotherapy.[5,6] Sifneos developed 'short-term anxiety-provoking psychotherapy' as a technique and theory with strict inclusion and exclusion criteria for choosing patients. Davanloo broadened the focus of the brief psychodynamic psychotherapies to include more than one conflict. He also expanded the inclusion criteria to individuals with character pathology and chronic phobic and obsessional neuroses, and advocated actively confronting resistances. Mann's time-limited psychotherapy identified a central issue related to the meaning of time, as the focus of the treatment. Mann related this to the patient's difficulties in confronting loss and separation and the reality of time and death.

In recent years, brief psychotherapy has become increasingly research based. Strupp, Luborsky, and Horowitz have all introduced manualized focused psychodynamic treatments which substantially contribute to our research understanding of this treatment modality.

Brief dynamic psychotherapy technique

(a) Evaluation and setting

The evaluation is particularly important in brief individual psychodynamic psychotherapy because of the need for rapid and accurate assessment. In contrast to longer term treatments, brief individual psychodynamic psychotherapy does not offer the luxury of time to re-evaluate and correct mistakes. Although at times we think of psychotherapy as beginning as soon as the doctor sees the patient, this is a hyperbole, used to underscore the importance of interpersonal and transferential elements in the initial meeting with the patient. In fact, it is extremely important, particularly in brief individual psychodynamic psychotherapy, to distinguish the diagnostic interviews from the ongoing treatment.

The interventions and technical procedures performed during the evaluation phase, usually one to four sessions, are substantially different from the technical aspects of brief individual psychodynamic psychotherapy itself. The evaluation phase includes the diagnosis, consideration of the interaction among the patient's ego strength, physical health, and selection variables, and the treatment recommendation, including considering the option that no treatment is indicated.

As in all medical treatments, brief individual psychodynamic psychotherapy is given to patients rather than to diseases. The ability to participate in brief individual psychodynamic psychotherapy process requires the patient to be able to access his or her fantasy life in an active and experiencing manner (i.e. psychologically minded) and, importantly, is able to get up and leave this process behind at the end of a session and not be lost in reverie or uncontrolled fantasies or fears. Note that this does not mean the patient requires a 'high IQ'. In fact, a high IQ, when accompanied with rigidity, intellectualization, and rumination, as is often seen, can be a contraindication to a brief psychodynamic treatment since these defences can be quite formidable. The availability of interpersonal support in the patient's real environment and the patient's ability to experience and simultaneously observe highly charged affective states are necessary to a successful treatment. Individuals who are in an emergent crisis (e.g. imminently suicidal, psychotic, recent major life trauma) and therefore are very concerned and focused on the real events in their life cannot enter into a brief psychodynamic psychotherapy without first having a period of supportive treatment. A true life crisis does not allow the patient the opportunity to explore fantasies.

Negotiation with the patient is an important part of reaching a treatment decision in brief individual psychodynamic psychotherapy. The patient must rapidly feel a part of the treatment and committed to the process. The process of setting a time limit at the beginning of the treatment can be an important element in decreasing the dropout rate from this form of treatment,[7] particularly with the patient who is concerned about dependency, 'becoming addicted' to the therapist, or who needs to maintain a substantial sense of control. What is dealt with in treatment can only be what the patient is able to bring into focus, what the patient can tolerate talking about, and what he or she can tolerate the therapist talking about.[8] Although this is not different than other psychodynamic treatments, the limited time of brief individual psychodynamic psychotherapy means that there is limited ability to interpret multiple defences that might open new areas of exploration.

(b) Technique

The rapid establishment of the therapeutic alliance is critical to brief individual psychodynamic psychotherapy.[9] Identifying the patient's initial anxieties related to beginning therapy is an important technique in the early sessions of brief individual psychodynamic psychotherapy in order to assure the alliance and to establish the conditions under which the patient can favourably hear and respond to the interpretations that the therapist will later give. As the therapy unfolds, the therapist operates on the hypothesis that each session is related to the previous one. The therapist strives in each session to identify the continuity of meaning related to the treatment focus that is present but hidden.[10] This continuity is driven by the 'experience bias' of the patient, and his or her tendency to experience the world in a certain way due to unique developmental experiences that have moulded his or her perception, interpersonal beliefs, and expectations.[11]

Brief individual psychodynamic psychotherapy is more focused, and more 'here and now' oriented with fewer attempts to reconstruct the developmental origins of conflicts than the extensive reworking of personality undertaken in longer term psychotherapies. Through the exploration of the patient's metaphors and symbols, both defensive patterns and disturbances in present interpersonal relations are identified in the treatment setting as well as in the patient's life. The importance of being able to hear what the patient has to say and to understand its meaning remains central as in other psychoanalytically oriented treatments.

Free association and inquiry: Free association is part of the technique of brief individual psychodynamic psychotherapy. But what constitutes free association—as in all dynamic therapies—requires thoughtful consideration. In its most basic form, and particularly highlighted in brief individual psychodynamic psychotherapy, free association means that the patient is free to choose what they wish to talk about. This rather direct definition emphasizes that free association is always relative. In addition, in brief dynamic psychotherapy, the patient is always somewhat more task focused that in open-ended treatments or psychoanalysis and this focus should not be discouraged by the therapist. Rather it is the therapist's task to hear the themes in the patient's concerns. The therapist asks questions, directs the patient's attention, and uses benign neglect, i.e. avoids some areas of conflict that cannot be dealt with at this time or in a short period of time. The therapist identifies those spots at which free association breaks down (the presence of a defence) or at which the narrative is carrying a single emotional story out of the patient's awareness. As in all dynamic treatments, often when the patient is able to talk freely and with a coherent narrative about their conflicts, the work of the treatment is completed.

Defence and transference: Brief individual psychodynamic psychotherapy emphasizes understanding (a) the mechanisms of defence used by the patient to decrease anxiety and other uncomfortable feelings associated with areas of conflict which are out of awareness, and (b) the characteristic transference relationships which distort the patients response to their adult world. Typically these two areas, defence and transference, create the world of meaning and expectations in which the patient lives. The techniques of the brief psychodynamic psychotherapy are directed towards clarifying these areas and presenting them to the patient to increase understanding and in this manner change symptoms and behaviour. Often only one defence is concentrated on in a given brief treatment. As the defence is clarified, the transference relationship may become evident. The developmental narrative of how the patient came to see the world in the way he or she does, provides the 'glue' through which the patient can integrate this knowledge into their life experience and behaviours, and recall it for practice and future use.

The brief individual psychodynamic psychotherapy therapist, similar to longer term psychodynamic work, must both enhance the patient's observing capacity in order that the transference can be observed by the patient and therapist, and create the therapeutic situation in which the patient can hear the therapist's interpretations in a useful manner. Dreams, as well as slips of the tongue and symptoms, can provide an avenue to the understanding of unconscious conflict which can be taught and explored with the

patient. The therapist strives to interpret both the triangle of anxiety (wish-defence-anxiety) and the triangle of insight (transference figure in the present—the therapist/patient interaction—transference figure from the past).

Frequently, when the transference is most evident, other elements of the past are simultaneously experienced in the patient's life. In brief individual psychodynamic psychotherapy these can be particularly important to the patient's understanding the feeling elements of the transference in a mutative manner since the depth and intensity of the transference is much less and much briefer than in long-term work. In contrast, however, the presence of a recent precipitant to the patient's problems, as is usually the case in brief psychodynamic psychotherapy, can considerably intensify transference responses and be a central element in developing the psychodynamic understanding for the patient. The transference experience—the transference, the life experiences being relived, and particularly the precipitant—provide the web of meaning that is the focus of interpretation and the mutative force in brief individual psychodynamic psychotherapy.

Often the transference in brief individual psychodynamic psychotherapy is paternal or maternal, but it has also been noted that, perhaps due to the time-limited nature of the work, sibling and transference figures from adolescence may more often be recalled in brief individual psychodynamic psychotherapy. The transference is rarely as deep as that seen in long-term treatment. It requires a skilled eye to note and bring the transference to the attention of the patient in a manner that is neither intrusive nor offensive.[12] Interpretations usually occur over several sessions, in the middle or later third of the treatment, during which past, present, and transference experiences are linked together. In the context of the affective arousal associated with this transference experience and the simultaneous understanding of the experience, behavioural change occurs and the patient's ability to perceive previously hidden feelings and relationships as well as his or her view of the future and the past can change.

Countertransference: Countertransference is also an important element in brief individual psychodynamic psychotherapy as in other psychodynamic treatments.[13] Analysis of countertransference reactions can allow the therapist to recognize subtle aspects of the transference relationship and to understand the patient's experience better. Because of the more active stance, the brief psychodynamic psychotherapist can be particularly prone to countertransferences that show up as over-involvement or aggression. In addition, the brief time available for treatment can make recovery from countertransference errors quite difficult.

(c) Medication

Medication is frequently used in conjunction with brief psychodynamic psychotherapy. This can complicate the treatment and its progress as well as aid in symptom recovery. The therapist must explore the meaning of the medication and its role in the patient's view of himself or herself and interpersonal strengths and vulnerabilities. At times, brief individual psychodynamic psychotherapy can also serve as an alternative to medication treatment for less severe symptoms or when medication is contraindicated. Medication may have also begun during the initial brief psychodynamic psychotherapy and then continued after the psychotherapy has formally stopped and the patient is followed with less frequent meetings to monitor medication. This sequence has many advantages including resolving present stressors and precipitants, encouraging medication compliance, and ongoing medical follow-up after therapy either in maintenance or intermittent frequency. Another course of brief dynamic therapy may be indicated at a later date if the response to combined treatment is ineffective or if new problems appear. Greater education of clinicians and research on this combined and sequential treatment is needed.

Comparison of the brief psychodynamic psychotherapies

The work of Malan, Sifneos, Mann, and Davanloo shows substantial overlap in each author's goals, selection criteria, technique, and duration of treatment.[14] The goals of all of these models of brief psychotherapy include facilitating health-seeking behaviours and mitigating obstacles to normal growth. From this perspective, brief psychotherapy focuses on the patient's continuous development throughout adult life and the context-dependent appearance of conflict, depending on environment, interpersonal relationships, biological health, and developmental stage. This picture of brief psychotherapy supports modest goals that require the therapist to refrain from perfectionism. Malan, Sifneos, Mann, and Davanloo also seem to agree with Stierlin's[15] contrast between brief psychotherapy's use of the 'propitious moment' and long-term treatment's use of 'a shared past' between therapist and patient. Both the propitious moment and the shared past carry psychotherapeutic advantages and disadvantages, emphasizing certain technical possibilities and limiting others.

Selection criteria: Many of the selection criteria emphasized by Malan, Sifneos, Mann, and Davanloo are common to all kinds of psychodynamic psychotherapy. However, unique selection criteria are required due to the brief duration of treatment. Patients in brief psychodynamic psychotherapy must be able to engage quickly with the therapist, terminate in a short period of time, and be able to carry on much of the working through and generalizing of the treatment effects on their own.

The necessity for greater independent action by the patient requires that the patient have high levels of ego strength, motivation, and responsiveness to interpretation. Sifneos's rather unique emphasis on intelligence as a criterion may be related to his anxiety-provoking interpretations, which require a broader educational context in order to be understood. The importance of the rapid establishment of the therapeutic alliance underlies a substantial number of the selection and exclusion criteria.

Focus of brief psychotherapy: All authors mention the central importance of the focus in brief psychotherapy, and therefore the evaluation sessions to determine this focus. Mann formulates the focus to the patient in terms of the patient's fears and pain. However, he would probably agree with Malan, Davanloo, and Sifneos in the importance of constructing the psychodynamic focus at a deeper level in one's own understanding of the work being done. Maintaining the focus is the primary task of the therapist. This enables the therapist to deal with complicated personality structures in a brief period of time. Resistance is limited through benign neglect of potentially troublesome but non-focal areas of the personality. The elaboration of techniques for establishing and maintaining the focus of treatment is critical to all brief individual psychodynamic psychotherapies.

Transference: The manner and rapidity in which transference is dealt with vary considerably among proponents of brief individual psychodynamic psychotherapy. Malan takes a more typical psychoanalytic approach of waiting for transference to become resistance before it is interpreted. Sifneos, in his emphasis on the Oedipal relationship, is more aggressive in handling the deep conflictual areas of transference material. Davanloo is confrontational in developing a transference experience. This confrontational style may at times confuse the patient's experience of the real and the transferential therapist. However, Davanloo often treats severe obsessional disorders. In these cases, the need to increase the patient's affective awareness is high. These may be the patients in which this particular technique is most useful. Aggressive, competitive, and hostile feelings, which might otherwise remain firmly defended, may thus become available to these patients.

Countertransference: The role of countertransference in brief psychotherapy is as complicated as it is in long-term treatment. Countertransference issues related to the aggressive techniques used by Sifneos and Davanloo have been observed. Countertransference experiences related to termination and loss can also be prominent.[16] The goal-directed techniques of brief psychotherapy limit the development of regressive countertransference responses.[13]

Duration of treatment: There is remarkable agreement on the duration of brief psychotherapy. Although the duration ranges from 5 to 40 sessions, authors generally favour 10 to 20 sessions. The duration of treatment is critically related to maintaining the focus within the brief psychotherapy. Shlien *et al.*[17] have found in Rogerian therapy, a correlation between the number of sessions and recovery. In general, they report an increasingly successful outcome (measured by the patient's self-concept) up to about 20 sessions. Howard *et al.*[18] using a meta-analytical technique, found 75 per cent of patients showing some improvement by 26 sessions. However, this study includes a wide range of types of treatment. When treatment extends beyond 20 sessions, the therapist frequently may find himself or herself enmeshed in a broad character analysis without a focal conflict. Change after 20 sessions may be quite slow. Clinical experience generally supports the idea that brief individual psychodynamic psychotherapy should be between 10 and 20 sessions although more complicated cases will require greater length of treatment. Often extending treatment beyond 20 sessions is recognition that treatment will be beyond 40 or 50 sessions.

Brief psychodynamic psychotherapy for depression, narcissistic disturbances, panic disorder, substance abuse, and post-traumatic stress disorder have been described.[14,19] Horowitz *et al.*[20] have described brief psychotherapy focused on the stress responses evidenced by various personality styles. He emphasizes that this psychotherapy is directed towards dealing with the process of the stress response and not character change. However, his outcomes indicate that selected character changes are possible in some areas. The distinction between recovery from a disruption in homoeostatic balance, reconstitution of self-esteem and self-concept, and changes in character structure require further exploration.

Critical points: The identification of critical points during brief psychotherapy, when the 'danger' of becoming a long-term treatment is most acute, clarifies the technical handling of brief psychodynamic psychotherapy. At these points, the therapist often notes an increasing vagueness of the goals of the treatment, decreased activity by the therapist, and the emergence of the transference as the central element. These variables indicate the potential of a short-term psychotherapy becoming a long-term treatment. The fourth to sixth hour of weekly 12-session therapy is often a point at which incipient or potential regression may suddenly appear. The patient at this time is testing the boundaries of the treatment. Action by the therapist is required if a brief psychotherapy is to remain exactly that—brief. The study of technical interventions, which occur at these critical moments, will further elucidate the technical handling of limited regression in brief psychodynamic psychotherapy.

Malan and the Tavistock group: focal psychotherapy: Developed from the workshops of Balint and Malan, focal psychotherapy is an example of applied psychoanalysis.[21] Malan has carried on Balint's earlier work.[22,23] Previous attempts to develop brief forms of psychoanalytic psychotherapy primarily involved the use of 'activity' which was frequently equated with manipulation. On the contrary, Malan emphasized the importance of choosing and maintaining a narrow focal area to be dealt with in a brief period of time. He stresses the importance of finding the appropriate focus in the patient's story and consistently interpreting the focal problem area.[23] Through selective attention and neglect, the therapist maintains the focus and completes a brief psychotherapy. The importance of determining the focus underscores the value of the diagnostic process, including the psychodynamic assessment of the patient prior to the initiation of psychotherapy.[24]

Malan identifies the following factors as leading to the lengthening of treatment: resistance, overdetermination, a need for working through the roots of conflict in early childhood, transference, dependence, negative transference connected with termination, and the transference neurosis. In addition, some therapist characteristics may lengthen treatment. These include a tendency towards passivity, a sense of timelessness conveyed to the patient, therapeutic perfectionism, and a preoccupation with deeper earlier experiences. All of these factors must be dealt with in order to maintain a brief therapy. For Malan, identifying a focal conflict acceptable to the patient is critical to a successful outcome (Table 6.3.4.1). In addition, the patient must have the capacity to think in feeling terms, demonstrate a high motivation, and exhibit a good response to trial interpretations made during the evaluation phase. Patients who have had serious suicidal attempts, drug addiction, long-term

Table 6.3.4.1 Brief psychodynamic psychotherapies

Goal of treatment
Identify the defence, the anxiety, and the impulse
Link the present, the past, and the transference
Focus of treatment
Internal conflict present since childhood
Selection criteria
Patient is able to think in feeling terms
Highly motivated
Good response to trial interpretation
Duration of treatment
Up to 1 year
Mean 20 sessions
Termination
Set definite termination date at beginning of treatment

hospital stays, more than one course of electroconvulsive therapy, chronic alcoholism, incapacitating severe chronic obsessional symptoms, severe chronic phobic symptoms, or gross destructive or self-destructive acting-out are excluded from treatment. The patient is also excluded from focal psychotherapy if the therapist anticipates any of the items in Table 6.3.4.2.

For Malan, the criteria in Table 6.3.4.2 represent specific dangers. If the therapist cannot make contact with the patient, or low motivation or rigid defences are present, it will be difficult to form an effective therapeutic working alliance within a short time. Complex or deep-seated issues, which must be dealt with to resolve a conflict area, require a longer period of treatment. Difficult transference relationships may also prevent timely termination or lead to premature termination. The occurrence of severe depressive or psychotic episodes during treatment can be a danger to the patient and require adjunctive treatments. Thus, Malan takes seriously the time limitation in brief therapy, which requires the rapid establishment of a therapeutic alliance and the ability to terminate therapy without the development of unexpected serious symptoms.

Malan, in contrast with other practitioners, does not automatically exclude patients with serious psychopathology. He sees the balance between motivation and focality as the primary criteria. A patient with only moderate motivation but a highly focal conflict might be accepted into treatment. Similarly, a patient with high motivation but not as focal a conflict might also be accepted into treatment with the hope that clarification of the focus would occur in a short period of time.

Identifying the precipitating factors, early traumatic experiences, or repetitive patterns can indicate the area of internal conflict present since childhood and the possible focus of treatment. The therapist should assess the congruence between the current conflict and the 'nuclear' or childhood conflict during the evaluation phase. The patient's response to interpretations about aspects of this conflict may lead to acceptance into treatment. According to Malan, the greater the probability that the conflict area will manifest itself in the transference, the more positive the outcome will be.

Malan is less concerned with technique than with the importance of choosing the focus. He employs the usual technical procedures of psychoanalytic psychotherapy and emphasizes the importance of making interpretations of the transference and connecting these to current and past relationships. This 'triangle of insight' (the transference, the current relationship, and the past relationship) leads to the patient's cure. Overall, the goal is to clarify the nature of the defence, the anxiety, and the impulse, which the patient is experiencing, and to link these to the present, the past, and the transference. Once the defence and the anxiety are clarified, the link to the past can be made. The interpretation that links to the past may be experienced as reassuring by the patient because of its emphasis on the conflict belonging to the world of fantasy rather than to the world of the present. Malan emphasizes transference interpretations as the most therapeutically effective interpretations because of their 'here and now' character.

In the brief therapy unit at the Tavistock Clinic, a time limit was almost always given at the beginning of treatment. For trainees this was usually 30 sessions. However, in his publications, Malan indicates a mean of 20 sessions for those cases with favourable outcomes. The longer time for trainees gives the opportunity to correct mistakes that might occur. In some published cases, therapy was extended up to 1 year (46 sessions). In general, Malan advocates the importance of a definite date rather than a number of sessions. Practically speaking, this eliminates the need for the patient and therapist to keep count of the number of sessions and eliminates complications related to whether or not to make up sessions that the patient has missed. Such a time limit gives a definite beginning, middle, and end to the therapy. It helps to concentrate the patient's material and the therapist's work, to maintain the focus, and decrease the diffuseness that might lead into long-term work.

Sifneos: short-term anxiety-provoking psychotherapy: Sifneos emphasizes the importance of patient selection because of the anxiety-provoking nature of his brief psychotherapy techniques (Table 6.3.4.1). He distinguishes anxiety-provoking therapy from anxiety-suppressing therapy, commonly referred to as supportive psychotherapy. For short-term anxiety-provoking psychotherapy, the patient must be of above average intelligence and have had at least one meaningful relationship with another person during his or her lifetime. The patient who has had such a relationship will be able to withstand the anxiety produced by the therapy and to develop rapidly a mature collaborative relationship with the therapist. This criterion tends to exclude narcissistic disorders. In addition, the patient must be highly motivated for change, not only for symptom relief. Sifneos also identifies several criteria for the patient selection based on the presentation of the patient during the evaluation. The patient must have a specific chief complaint. If the patient has a number of complaints, Sifneos asks the patient which complaint is of top priority. The patient's ability to identify one conflict area and to postpone work on others is taken as an indication of the patient's ability to tolerate anxiety. Sifneos looks for patients with anxiety, depression, phobias, conversion, and mild obsessive-compulsive features or personality disorders involving clear-cut interpersonal difficulties. During the evaluation, the patient must show an ability to interact with the evaluating psychiatrist, to express feelings, and to show some flexibility.

Sifneos is one of the few authors who clarifies his assessment of motivation. He defines motivation as including the patient's ability to recognize symptoms as psychological, a tendency to be introspective and honest about emotional difficulties, and a willingness to participate in the treatment situation. In addition, motivation includes curiosity, willingness to change as well as a willingness to make reasonable sacrifices, and a realistic expectation of the results of psychotherapy.

Sifneos focuses on the Oedipal conflict and does not expect a good outcome in dealing with other than Oedipal conflict areas.

Table 6.3.4.2 Exclusion criteria for Malan's and the Tavistock group's brief focal psychotherapy

1. Therapist is unable to make affective contact with the patient during the evaluation
2. Therapist anticipates that extended work will be needed
   To generate motivation
   To decrease rigid defences
   To reach complex or deep-seated issues
   To resolve unfavourable, intense transference, or dependence which may develop
3. Depressive or psychotic disturbance may intensify and place the patient at risk

The majority of failures using short-term anxiety-provoking psychotherapy have occurred in patients who complained of reactive depression following the loss of a loved one. He believes that this failure is due to the non-triangular (non-Oedipal) origins of the ambivalent feelings in some patients. In such cases, when the issue of termination arises, the patient regresses and an impasse is reached.

During the initial phase of psychotherapy, the therapist must establish good rapport with the patient in order to create a therapeutic alliance. The therapist uses anxiety-provoking confrontations in order to clarify issues around the patient's early life situation and present-day conflict. The therapist avoids areas such as passivity, dependence, and acting-out, which might lead to extensive regression. The use of anxiety-provoking confrontations in a direct attack on the patient's defences distinguishes short-term anxiety-provoking psychotherapy from other brief psychotherapies. Although it is made clear to patients during their evaluation that the psychotherapy is expected to last only a few months, no specific number of sessions or termination date is given. Interviews are held weekly and last for 45 min. The vast majority of treatments last from 12 to 16 sessions, and none go beyond 20 sessions. The aggressive confrontational style of this treatment underscores the importance of excluding pre-Oedipal problems and the importance of countertransference reactions in the therapist related to being too aggressive.

Mann: time-limited psychotherapy: Mann has focused on the specific limitation of time in brief psychotherapy. Mann sees the variable of time as a specific operative factor in psychotherapy as well as an element in its curative effect.[25,26] The experiences of the timelessness of treatment and of the treatment's termination are significant elements in Mann's view of the psychotherapeutic process.

Usually there are two to four evaluation meetings prior to beginning psychotherapy. Mann limits psychotherapy to a total of 12 treatment hours, distributed according to patient need. This may result in weekly 30-min sessions for 24 weeks or twice weekly hour-long sessions for 6 weeks. In practice, however, nearly all patients are seen in once-weekly 45- or 50-min sessions for 12 weeks. Mann admits having chosen the number 12 somewhat arbitrarily; however, his clinical experience indicates that somewhere between 10 and 14 sessions is a sufficient number. Mann emphasizes the importance of a uniform number of sessions for evaluating the psychotherapeutic process among different therapists. In this way, the relationship between the patient's presenting problems and psychotherapeutic technique can be more easily studied. Also, the provision of a specific number of sessions can be more easily accepted by the patient as a typical medical 'prescription'. Finally, the setting of a specific last session in the initial contract with the patient allows the therapy to have a clear beginning, middle, and end (see Table 6.3.4.1).

Mann indicates a number of exclusionary criteria: serious depression, acute psychosis, borderline personality organization, and the inability to identify a central issue. Mann sees Sifneos' criteria as primarily excluding borderline patients. He does not agree with Sifneos' emphasis on superior academic or work performance.

To some extent, Mann initially minimized selection as a central issue for brief psychotherapy. Later, Mann expanded his selection criteria by emphasizing the importance of the patient's ego strength as measured by prior work performance and past relationships.[26] Patients who may have difficulty engaging and disengaging rapidly from treatment are excluded. This includes schizoid patients, certain obsessional patients, patients with strong dependency needs, some narcissistic patients, some depressive patients who will not be able to form a rapid therapeutic alliance, and some patients with psychosomatic disorders who do not tolerate loss well.

According to Mann, the selection of the central issue for the psychotherapy is the critical event. It is the vehicle through which the patient is engaged in the work of therapy and on which a successful outcome depends. Mann looks for a central issue that is developmentally and adaptively relevant and has been recurrent over time. He describes this issue as the patient's 'present and chronically endured pain' and characterizes it as preconscious. Mann has further described the central issue as including a particular image of the self.[25] The central issue formulated in terms of time, affect, and an image of the self is the 'paradigm of the transference' expected to emerge in treatment. The therapist's statement of the central issue is a clarification, which can be readily recognized, felt, and held onto by the patient. Time-limited psychotherapy is intended to resolve this present and chronically endured pain and the patient's 'negative self-image'. The therapist frames the central issue to the patient in terms of a general statement about feelings.

Mann and Goldman[26] described in detail the phrasing of the central issue to the patient. It is the central issue that specifies the therapeutic contract and the goal of the therapy. In the case of a 41-year-old depressed woman who was preoccupied with her husband and children being even a minute late, Mann suggested the central issue: 'You've encountered extreme life situations and have managed them remarkably well ... yet you fear and have always feared that despite your best efforts you will lose everything'. In a 31-year-old married man attempting to gain a college degree who was consumed with a fear of failing, Mann suggested the central issue: 'Because there have been a number of sudden and very painful events in your life, things always seem uncertain, and you are excessively nervous because you do not expect anything to go along well. Things are always uncertain for you'.[26]

Mann uses the usual psychoanalytic psychotherapy techniques: defence analysis, transference interpretation, and genetic reconstruction. Transference is interpreted from within the central identified conflict area and in terms of the adaptive processes of the patient. However, Mann does not confront the patient. In general, his interventions are very close to the conscious material provided by the patient. Mann identifies specific dynamic events that unfold during the 12 sessions. The opening sessions are understood as filled with the unconscious magical expectation that past pains will now be resolved. During the initial phase, the therapist makes few comments and accepts the positive transference of the patient. Important aspects of the current problem, defence mechanisms, coping styles, and genetic roots of the central issue become clearer during this phase. In the middle four sessions, resistance is likely to appear, as well as the negative transference. The patient experiences the frustration that all of the wished for changes may not occur. In the ending phase of treatment, termination and the patient's resistances to termination in the face of unresolved problems in other areas of life are prominent.

Mann sees the importance of confronting separation and termination issues as critical to the success of brief psychotherapy. Frequently, the patient unconsciously reveals an awareness that the mid-point of treatment has come. The patient experiences

separation from the transference-invested therapist as a separation from an ambivalently experienced person from the past, without having achieved the fantasized magical resolution. The goal is to enable the patient to separate from the transference-invested therapist less ambivalently than he had done from this earlier important figure. Consequently, both the resolution of the central issue and the unfolding of an attachment-separation process in the 12-session treatment contract are intimately related through the development and interpretation of the transference.

Davanloo: broad-focus short-term dynamic psychotherapy: Davanloo writes about broad-focus short-term dynamic psychotherapy.[27] His selection criteria include patients with an Oedipal focus, those with a loss focus, and those with multiple foci. Davanloo is particularly interested in patients suffering from long-standing obsessional and phobic neuroses. His research data indicate that 30 to 35 per cent of the psychiatric outpatient population can benefit from this mode of therapy. Most information about his technique is derived from the publication of cases, presentations, and brief descriptions of his research that accompany case presentations.

The initial evaluation is a specific focused interview in which the patient's defences against 'true' feelings are gently but consistently confronted. Davanloo says that this is not a universal technique for the initial interview and cautions on its use with patients with severe psychopathology. Selection is based on psychological mindedness, the quality of the patient's interpersonal relations, and, in particular, on the presence of at least one meaningful relationship in the patient's past. The patient's ability to tolerate and experience anxiety, guilt, and depression are important (Table 6.3.4.1). The patient must be motivated to complete the treatment process and to resolve neurotic problems. His or her ability to respond to interpretation is an important selection criterion. In particular, response to transference interpretations, which link the transference with the present and the past, is a critical feature in the assessment for broad-focus short-term dynamic psychotherapy. Davanloo finds no value in criteria based on severity and duration of illness. Finally, the presence of flexibility in the ego's defensive pattern and a lack of use of the primitive defences of projection, splitting, and denial are important factors in selecting patients.

The technique Davanloo uses in therapy is a continuation of that used in the initial interview. The emotional experience of the patient in the transference is emphasized. The patient is 'gently but relentlessly' confronted about his defences against feelings in the transference relationship and in the past. All the usual techniques of psychoanalytic psychotherapy are employed: defence analysis, transference interpretations, and genetic reconstruction. Dreams and fantasy materials are also used. Transference interpretations tend to be made early. Because of the confrontive style, a strong therapeutic alliance is necessary. Patients frequently experience hostile, angry feelings towards the therapist because of being confronted. Davanloo actively pursues the patient's defences against recognizing the anger and its transference elements. Davanloo warns therapists that passive dependent and obsessional characters may develop a symbiotic transference relationship. This may be avoided through active confrontation and selection of patients. The active confrontation of defences and early transference interpretations tend to mobilize powerful affects and memories early on in treatment.

Davanloo recommends from 5 to 40 sessions, depending on the patient's conflict area (Oedipal versus multiple foci) and other selection criteria. In general, his treatments fall between 15 and 25 sessions. He does not recommend setting a specific termination date but rather makes clear to the patient that treatment will be short. Shorter time periods (5–15 sessions) are chosen for patients with a predominantly Oedipal focus, longer durations (20–40 sessions) for the more seriously ill group.

Comparison of psychodynamic, cognitive, and interpersonal brief psychotherapies

Interpersonal psychotherapy[28,29] and cognitive behavioural psychotherapy[30] derive from the psychodynamic model and therefore share many common elements with brief psychodynamic psychotherapy but with distinct approaches and interventions. All three modalities, interpersonal psychotherapy, cognitive behavioural therapy, and brief individual psychodynamic psychotherapy, are complex methods of treatment that must be custom-tailored to the individual patient. Brief by definition, they all lack the extended working through and application period of psychoanalysis and intensive (long-term) psychodynamic psychotherapy. All demand a high degree of clinical judgement and considerable experience to acquire competency. The relationship between the therapist and patient and the establishment of a therapeutic alliance are essential (Table 6.3.4.3).

Table 6.3.4.3 Comparison of the brief dynamic psychotherapy with cognitive psychotherapy and interpersonal psychotherapy

	Brief dynamic psychotherapy	Cognitive psychotherapy	Interpersonal psychotherapy
Free association	++	+	+
Directiveness	+	+++	++
Neutrality	+++	+++	+++
Time-limited	+++	+++	+++
Defence analysis	+++	+++ Schema/distortions	+
Transference	+++ Interpersonal	+	+++ patterns
Behavioural interventions	—	+++	+
Published manuals	+	++	++
Concurrent use of medication	++	+++	+++
Empirical research indicates efficacious treatments	+	+++	+++
Training in long-term dynamic psychotherapy helpful	+++	+	+++

While sharing many similarities, it is ultimately in the conception of the problem, the goals, and therapeutic interventions that these treatments differ. It is unclear to what extent behavioural changes may be attributed to the similarities or differences between treatments. All psychotherapies, including brief individual psychodynamic psychotherapy, interpersonal psychotherapy, and cognitive behavioural therapy teach new skills-problem-solving skills directed at how to resolve interpersonal and emotional problems when they arise. Differences among these psychotherapies in their interventions are more striking than the differences in their goals or the problem areas they identify for therapeutic work. In psychodynamic psychotherapy the structure of the session is determined by the flow of the patient's thoughts and their interaction with the therapist's interpretive comments. In contrast, cognitive and interpersonal psychotherapies use more directive, structured, and behavioural interventions. Whereas the brief individual psychodynamic psychotherapy like other psychodynamic psychotherapies relies on the patient to activate and practice new behaviours without direction. The therapist remains an empathic interpreter, a sharer of the patient's experience and perspective. While in other therapies, especially cognitive, the therapist may direct, prescribe, enjoin, educate, or role play.

Practical problems in brief psychodynamic psychotherapy

The choice of focus is perhaps the most important and the most difficult aspect of brief individual psychodynamic psychotherapy. It is helpful to identify several foci during the evaluation process, recognizing that there are inevitably several conflict areas active at any one time in a patient's life. Then the therapist can begin the process of thinking through what the treatment of each focus would entail (Table 6.3.4.4).

The therapist can begin to decide which focal conflict will be more difficult to reach in a brief period of time, which will threaten the therapeutic alliance more and therefore require a deeper working relationship that may take more time, and which focus requires interpreting more primitive defences and therefore may be more complicated.

Choice of a particular focus can also create more family or external disruption or support which can aid or disrupt the treatment.

Use of medication requires carefully explaining to the patient the relationship of the medication to the psychotherapy. Often the medication treatment will continue beyond the psychotherapy.

Table 6.3.4.4 Identifying and selecting the focal conflict in brief dynamic psychotherapy

Identifying the focal conflict
Explore
Precipitant of symptoms
Early life traumas
Repetitive patterns of behaviour
Listen for inhibitions/avoidance
Watch for conflicts about success as well as loss/failure
Selection among several foci
Choose the focus that is presently active
Use trial interpretation to identify active focus
Select focus related to only one transference figure

If repeated complicated medication alterations are needed or if serious side effects of the medication occur, the psychotherapy plan may have to be altered to allow time to understand them from the patient's perspective.

New therapists are often concerned about setting the date of termination at the time of the evaluation, fearing that they may not be able to complete the work by the deadline. Supervision with an experienced colleague can be very helpful to assure confidence and avoid mistakes that may lengthen the treatment. Alternatively, the new therapist may feel too much relief in setting the termination date when treating a very dependent patient and therefore miss the intensity with which the patient is attached and experiencing the therapist as an important, needed, or feared figure from the past.

The management of missed sessions should be made clear at the beginning of treatment. Usually it is best not to 'make up' the sessions, but to keep to the termination date. If the therapist is concerned about this as a potential issue in the treatment, the therapist may wish to plan several additional sessions in the overall treatment to assure this can be discussed and understood therapeutically. Of course if an emergency arises it is always appropriate to schedule appointments as needed for the health and safety of the patient.

The patient who 'divulges' new 'secret' information near the end of the treatment is a challenge to all therapists. Understanding to what extent this represents narcissistic, or sociopathic issues, fear of the therapist or the treatment, or the emergence of hope for the future or a transference enactment will determine how to respond.

Brief individual psychodynamic psychotherapy is best learned in conjunction with the skills of longer term psychodynamic psychotherapy. In the longer work, the therapist will be able to see more easily the possible conflict areas and think about the sequencing of the treatment of these, i.e. which is closer to the patient's awareness or which is more defended. In addition there is more time to correct errors and repair untoward events in the therapeutic relationship. The brief individual psychodynamic psychotherapist will have less time to correct mistakes and must more quickly identify conflict areas and assess their relative importance and potential for resolution through treatment.

Efficacy: research and evaluation

The brief psychodynamic treatments have a small empirical database. Much further research is needed.[31] In general, studies have supported the efficacy of this treatment approach. However, methodological issues are prominent in most research in this area. The development of handbooks for treatment has gone far in improving research in the psychoanalytically oriented brief treatments.[19,32–34]

The effectiveness of psychotherapy in general, is not argued as in the past.[8,35–37] Brief psychodynamic psychotherapy has been shown to have an effect size similar to many other medical treatments. Short-term psychodynamic psychotherapy has shown modest to moderate, often sustained gains for a variety of patients.[38] However, the question of which psychotherapy is suitable for which patient and by which therapist is still unclear. The cost-effectiveness of psychotherapeutic treatment remains hotly debated and is a focus of substantial research.[9,39,40] Individual psychotherapy has been shown to result in fewer days of hospital stay for patients on medical or surgical services of a general hospital. In health clinics

or health maintenance organizations, brief psychotherapy decreases the number of visits to primary health care providers, reduces the number of laboratory and radiographic studies, decreases the number of prescriptions given, and, overall, reduces direct health care costs. Recently summaries of the cost-offset effects of outpatient mental health treatment, the majority of which were short-term are hopeful but not unambiguous. One study found outpatient psychotherapy resulted in a 33 per cent average reduction in medical care utilization. Furthermore, these reductions occurred mostly in the more expensive, inpatient medical services. In another study, 72 patients with significant emotional problems and treated only by internists in a general medical clinic were compared with 62 patients who, in addition to being treated by internists for medical problems, received 10 weekly psychotherapy visits. Both groups had approximately an equal degree of emotional disturbance. At 4-month and 1-year follow-ups, the brief psychotherapy group reported significantly more global improvement than the non-psychotherapy group. Also, more patients in the brief psychotherapy group became employed at 1-year follow-up than in the non-psychotherapy group. This study suggests specific beneficial effects of brief psychotherapy when used in a medical setting by skilled psychotherapists. Combining psychotherapy with antidepressant medication has also been shown to give the best outcome at 1 year when compared to either treatment alone. Whether a therapist keeps to a consistent frame of reference in the treatment may also be a predictor of success if brief individual psychodynamic psychotherapy, regardless of what that perspective is.[41]

Malan's finding of the importance of making the transference-parent link for the successful outcome of treatment is significant and requires further exploration.[2] One reanalysis of Malan's data confirmed his finding and one did not.[42] In addition, one replication of this finding has been published.[43] Importantly, more recently, the overuse of transference interpretations has been shown to lead to poorer outcome. The therapeutic alliance, particularly when measured from the patient's perspective, has a consistent although modest contribution to outcome.[9,44] It has been shown that independent of the type of treatment and early clinical improvement, the therapeutic relationship contributes directly to the positive therapeutic outcome.[45]

The quality of the therapeutic interaction and the handling of the transference and countertransference appears to be critical to success or failure in brief individual psychodynamic psychotherapy.[34] Patients treated by therapists who have not been professionally trained, may, on average, be as improved as patients treated by professional brief dynamic therapists. However, such non-experienced therapists run out of relevant material and are unwilling to continue to treat patients over an extended period of time.[46] One of the important tasks of training in psychotherapy may be the development of the ability to 'endure' with the patient and, over time, with numbers of patients. Technical training and a theoretical framework may allow the therapist to maintain a sense of competence, direction, and interest in the work which the non-professional therapist cannot.

Interpersonal psychotherapy and cognitive-behavioural therapy have been much more extensively studied than dynamic psychotherapy, particularly in combination with medications. Recently, telephone psychotherapy with cognitive behavioural therapy in primary care settings when initiating antidepressant medication has been shown to improve clinical outcome.[47,48] To the extent that these treatments share techniques and outcomes, similar results might be expected with brief dynamic psychotherapy; however, this still needs to be shown. Focal directive psychotherapies generally appear to be more effective than traditional unstructured psychodynamic psychotherapy for a number of types of patients, but a delineation of which psychotherapy for which patient over what time and with which medication remains to be demonstrated. Good clinical sense dictates combined treatments with matching the patient's cognitive and affective style with treatment type and making medication compliance a focus of any psychotherapy. Additionally, further research of brief psychodynamic psychotherapies in specific psychiatric disorders as well as across the life span are needed.[49,50]

Conclusion

Brief dynamic psychotherapy is an important treatment for numerous disorders, primarily the adjustment, anxiety, and mood disorders. Both alone and in combination with medication brief dynamic psychotherapy is an effective part of the treatment armamentarium. Clinicians should be trained in the brief as well as the longer term treatments and their use as brief, intermittent, and maintenance treatments. Skill in the longer term psychotherapies is important to developing skill in the brief dynamic psychotherapy where the needs for rapid establishment of the therapeutic alliance and the accurate assessment of transference and defence patterns are important.

Empirical studies comparing well-defined brief dynamic psychotherapy with cognitive and interpersonal psychotherapies are limited. Future research must address which form of brief psychotherapy may be most helpful for which patient. An individual's preferred learning path-what he or she may see and observe most easily such as thoughts or feelings or interpersonal relations-may be an important variable in determining which brief psychotherapy for which patient. State, trait, and contextual variables will influence this learning modality. The process of change in brief individual psychodynamic psychotherapy, a process of altering neuronal organization through verbal means, is influenced by the patient's diagnosis, medications, past history, cognitive style, developmental stage, and affective availability, as well as the doctor–patient match.

Further information

Ursano, R.J., Sonnenberg, S., and Lazar, S. (2004). *Concise guide to psychodynamic psychotherapy: principles and techniques of brief, intermittent and long term psychodynamic psychotherapy*. American Psychiatric Press, Washington, DC.

Levinson, H., Butler, S.F., Powers, T.A., *et al.* (2002). *Concise guide to brief dynamic and interpersonal psychotherapy* (2nd edn). American Psychiatric Publishing Inc., Washington, DC.

Luborsky, L. and Luborsky, E. (2006). *Research and psychotherapy: the vital link*. Jason Aronson, Lanham MD.

Dewan, M.J., Steenbarger, B.N., and Greenberg, R.P. (eds.) (2004). *The art and science of brief psychotherapies: a practicioner's guide*. American Psychiatric Publishing Inc., Arlington, VA.

References

1. Gabbard, G.O. (1994). *Mind and brain in psychiatric treatment*. Institute of Pennsylvania Hospital Strecker Award Monograph Series 31. Pennsylvania Hospital, Philadelphia, PA.

2. Ursano, R.J. and Silberman, E.K. (1988). Individual psychotherapies. In *Textbook of psychiatry* (eds. J.A. Talbott, R.E. Hales, and S.C. Yudofsky), pp. 855–89. American Psychiatric Press, Washington, DC.
3. Michels, R. (1997). Psychodynamic psychotherapy in modern psychiatry. *Journal of Practical Psychiatry and Behavioral Health*, **3**, 95–8.
4. Ferenczi, S. and Rank, O. (1925). *The development of psychoanalysis*. Nervous and Mental Diseases Publishing Company, New York.
5. Sifneos, P.E. (1972). *Short-term psychotherapy and emotional crisis*. Harvard University Press, Cambridge, MA.
6. Sifneos, P.E. (1984). The current status of individual short-term dynamic psychotherapy and its future: an overview. *American Journal of Psychotherapy*, **37**, 472–83.
7. Sledge, W.H., Moras, K., Hartley, D., *et al.* (1990). Effect of time-limited psychotherapy on patient dropout rates. *The American Journal of Psychiatry*, **147**, 1341–7.
8. Crits–Christoph, P. and Barber, J.P. (eds.) (1991). *Handbook of short-term dynamic psychotherapy*. Spectrum, New York.
9. Ursano, A.M., Sonnenberg, S.M., and Ursano, R.J. (in press). Physician patient relationship in psychiatry (3rd edn) (eds. A. Tasman, J. Kay, and J.A. Lieberman). Wiley & Sons, Ltd., West Sussex, England.
10. Coleman, J.V. (1968). Aims and conduct of psychotherapy. *Archives of General Psychiatry*, **18**, 1–6.
11. McGuire, M. (1965). The process of short-term insight psychotherapy. *The Journal of Nervous and Mental Disease*, **141**, 83–94.
12. Frances, A. and Perry, S. (1983). Transference interpretations in focal therapy. *The American Journal of Psychiatry*, **140**, 405–9.
13. Klan, H. and Frances, A. (1984). Countertransference in focal psychotherapy. *Psychotherapy and Psychosomatics*, **41**, 38–41.
14. Levinson, H., Butler, S.F., Powers, T.A., *et al.* (2002). *Concise guide to brief dynamic and interpersonal psychotherapy* (2nd edn). American Psychiatric Press, Washington, DC.
15. Stierlin, H. (1968). Short-term versus long–term psychotherapy in the light of a general theory of human relationships. *British Journal Medical Psychology*, **41**, 357–67.
16. Hoyt, M. and Farrell, D. (1984). Countertransference difficulties in a time–limited psychotherapy. *International Journal of Psychoanalytic Psychotherapy*, **10**, 191–203.
17. Shlien, J.M., Mosik, H.H., and Dreikurs, R. (1962). Effective time limits: a comparison to psychotherapy. *Journal of Counselling Psychology*, **9**, 31–4.
18. Howard, K., Kopta, S., Krause, M., *et al.* (1986). The dose-effect relationship in psychotherapy. *The American Psychologist*, **41**, 159–64.
19. Milrod, B.L., Busch, F.N., and Cooper, A.M. (eds.) (1996). *Manual of panic-focused psychodynamic psychotherapy*. American Psychiatric Press, Washington, DC.
20. Horowitz, M.J., Marmar, C., Krupnick, J., *et al.* (1984). *Personality styles in brief psychotherapy*. Basic Books, New York.
21. Balint, M., Ornstein, P., and Balint, E. (1972). *Focal psychotherapy*. Lippincott, Philadelphia, PA.
22. Malan, D.H. (1975). *A study of brief psychotherapy*. Plenum Press, New York.
23. Malan, D.H. (1980). *Toward the validation of dynamic psychotherapy*. Plenum Press, New York.
24. Ursano, R.J., Sonnenberg, S., and Lazar, S. (2004). *Concise guide to psychodynamic psychotherapy: principles and techniques of brief, intermittent and long term psychodynamic psychotherapy.* American Psychiatric Press, Washington, DC.
25. Mann, J. (1980). *Time–limited psychotherapy*. Harvard University Press, Cambridge, MA.
26. Mann, J. and Goldman, R. (1995). *A casebook in time-limited psychotherapy*. Jason Aronson, New York.
27. Davanloo, H. (ed.) (1980). *Short-term dynamic psychotherapy*. Jason Aronson, New York.
28. Markowitz, J.C. (ed.) (1998). *Interpersonal psychotherapy*. American Psychiatric Press, Washington, DC.
29. Klerman, G.L., Weissman, M.M., Rounsaville, B.J., *et al.* (1984). *Interpersonal psychotherapy of depression*. Basic Books, New York.
30. Beck, A.T. and Rush, A.J. (1995). Cognitive therapy. In *Comprehensive textbook of psychiatry*, Vol. VI (eds. H.I. Kaplan and B.J. Sadock), pp. 1847–57. Williams and Wilkins, Baltimore, MD.
31. Kay, J. (1997). Brief psychodynamic psychotherapies: past, present and future challenges. *Journal of Psychotherapy Practice and Research*, **6**, 330–7.
32. Luborsky, L. (2000). *Principles of psychoanalytic psychotherapy: a manual for supportive expressive treatment*. Basic Books, New York.
33. Miller, N.E., Luborsky, L., Barber, J.P., *et al.* (eds.) (1993). *Psychodynamic treatment research*. Basic Books, New York.
34. Strupp, H.H. and Binder, J. (1984). *Psychotherapy in a new key: time-limited dynamic psychotherapy*. Basic Books, New York.
35. Gabbard, G.O., Lazar, S.G., Hornberger, J., *et al.* (1997). The economic impact of psychotherapy: a review. *The American Journal of Psychiatry*, **154**, 147–55.
36. Lazar, S.G. (ed.) (1997). *Extended dynamic psychotherapy: making the case in an era of managed care.* Psychoanalytic inquiry supplement. Analytic Press, Hillsdale, NJ.
37. Crits–Christoph, P. (1992). The efficacy of brief dynamic psychotherapy: a meta-analysis. *The American Journal of Psychiatry*, **149**, 151–8.
38. Abass, A.A., Hancock, J.T., Henderson, J., *et al.* (2007). *Short-term psychodynamic psychotherapies for common mental disorders* (*Review*), pp. 1–47. The Cochrane Collaboration, John Wiley & Sons, Ltd, New York.
39. Ursano, R.J. and Silberman, E.K. (1999). Psychoanalysis, psychoanalytic, psychotherapy, and supportive psychotherapy. In *Textbook of psychiatry* (eds. R.E. Hales, S.C. Yudofsky, and J.A. Talbot), pp. 1157–84. American Psychiatric Press, Washington, DC.
40. Wiborg, I.M. and Dahl, A.A. (1996). Does brief dynamic psychotherapy reduce the relapse rate of panic disorder? *Archives of General Psychiatry*, **53**, 689–94.
41. Barbar, J.P., Crits–Christoph, P., and Luborsky, L. (1996). Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of Consulting and Clinical Psychology*, **64**, 619–22.
42. Hoglend, P. (1996). Long-term effects of transference interpretations: comparing results from a quasi-experimental and a naturalistic long-term follow-up study of brief psychotherapy. *Acta Psychiatrica Scandinavica*, **93**, 205–11.
43. Marziali, E.A. (1984). Prediction of outcome of brief psychotherapy from therapist interpretive interventions. *Archives of General Psychiatry*, **41**, 301–4.
44. Joyce, A.S., Ogrodniczuk, J.S., Piper, W.E., *et al.* (2003). The alliance as mediator of expectancy effects in short-term individual therapy. *Journal of Consulting and Clinical Psychology*, **71**, 672–9.
45. Zuroff, D.C. and Blatt, S.J. (2006). The therapeutic relationship in the brief treatment of depression: contributions to clinical improvement and enhanced adaptive capacities. *Journal of Clinical Psychology*, **74**, 130–40.
46. Strupp, H.H. (1980). Success and failure in time-limited psychotherapy: with special reference to the performance of lay counselors. *Archives of General Psychiatry*, **37**, 831–41.
47. Simon, G.E., Ludman, E.J., Tutty, S., *et al.* (2004). Telephone psychotherapy and telephone care management for primary care patients starting antidepressant treatment: a randomized controlled trial. *The Journal of the American Medical Association*, **292**, 935–42.
48. Ludman, E.J., Simon, G.E., Tutty, S., *et al.* (2007). A randomized trial of telephone psychotherapy and pharmacotherapy for depression: continuation and durability of effects. *Journal of Consulting and Clinical Psychology*, **75**, 257–66.

49. Leichenring, F, Rabung, S, and Leibing, E. (2004). The efficacy of short-term psychodynamic psychotherapy in specific psychiatric disorders. *Archives of General Psychiatry*, **61**, 1208–16.
50. Shefler, G. (2000). Time-limited psychotherapy with adolescents. *Journal of Psychotherapy Practice and Research*, **9**, 2.

6.3.5 Psychoanalysis and other long-term dynamic psychotherapies

Peter Fonagy and Horst Kächele

Introduction

Basic assumptions

The term psychodynamic psychotherapy has no specific referent. It denotes a very heterogeneous range of psychological treatment approaches which arguably have in common an intellectual heritage of psychoanalytic theory. Psychoanalytic theory itself is no longer based on a unitary body of ideas[1] but a number of ideas appear to be core to most psychodynamic approaches. These notions are:

(a) A shared notion of psychological causation, that mental disorders can be meaningfully conceived of as specific organizations of an individual's conscious or unconscious beliefs, thoughts, and feelings.

(b) Psychological causation extends to the non-conscious part of the mind, and to understand conscious experiences, we need to refer to other mental states of which the individual is unaware.

(c) The mind is organized to avoid unpleasure arising out of conflict[2] in order to maximize a subjective sense of safety.[3]

(d) Defensive strategies are a class of mental operations that seem to distort mental states to reduce their capacity to generate anxiety, distress, or displeasure. Individual differences in the predisposition to specific strategies have often been used as a method for categorizing individuals or mental disorders.[4,5]

(e) Varying assumptions are made concerning normal and abnormal child and adolescent development but therapists are invariably oriented to the developmental aspects of their patients' presenting problems.[6]

(f) Relationship representations linked with childhood experience are assumed to influence interpersonal social expectations including the transference relationship with the therapist[7] and to shape the representations of the self.[8–11]

(g) These relationship representations inevitably re-emerge in the course of psychodynamic treatments.[12]

Brief overview of theories

Psychoanalytic theory has evolved from the work of Freud following two broadly separate paths which converged over the past 25 years only to separate again. In the United States followers of the Vienna school in the 1950s and 1960s evolved a systematic psychology of the ego, a conflict-oriented complex psychological model of the mind and its disturbances.[13] In Europe, only Anna Freud and her followers in London pursued this tradition of psychoanalytic thought.[14] Based on the Berlin school of Karl Abraham, Melanie Klein and her followers established a distinct approach focusing on the understanding of disturbance rooted in infantile destructiveness and sadism.[15] Some psychoanalysts, influenced by Klein and the idea of the pathogenic nature of the experiences of infancy, gradually discarded the mechanistic psychology of drives and psychology of internal structures in favour of theories of intrapsychic interpersonal relationships (object-relations theory).[16]

As these schools developed in the United Kingdom, their influence travelled across the Atlantic. First, Kohut, strongly influenced by Winnicott (albeit without explicit acknowledgement), evolved a psychoanalytic psychology of the self.[17] Shortly after, Kernberg arrived at an imaginative integration of ego-psychological and Kleinian ideas.[18] In the meantime, in the United Kingdom, the Kleinian movement rapidly progressed in their understanding of psychoanalytic clinical experience, moving beyond Klein's original work and integrating some of the key features of the Anna Freudian and the British object-relations traditions.[19] In the United States, disillusionment with the false certainty provided by ego-psychology became intense throughout the late 1970s and early 1980s and a radical change in psychoanalytic thinking took place with the emergence of the interpersonal relational perspective, which is in part rooted in the work of Harry Stack Sullivan.[20,21] The relational psychoanalysis of the 1980s and 1990s consolidated several lines of thought initiated by justified critiques of traditional analytic theory[22]; including feminism, the hermeneutic-constructivist critique of the analyst's authority, infancy research, and, closely related to this, the intersubjectivist-phenomenological philosophy of mind—as well as a general political movement to improve and democratize access to analytic ideas and training.[23]

There are many other new psychoanalytic theoretical approaches, bringing the field increasingly close to total fragmentation.[24] This is because the emergence of new approaches in no way signals the demise of any previous orientations, most of which continue to enjoy considerable popularity among specific groups of psychoanalysts.

Psychoanalytic therapy as treatment

The history of psychoanalysis as a therapeutic approach is rather different. Broadly speaking, it may be argued that psychoanalysis and other long-term psychodynamic therapies are predominantly verbal, interpretive, insight-oriented approaches which aim to modify or re-structure maladaptive relationship representations. It is implicitly assumed that genetic and early environmental factors give rise to partial, unintegrated, and generally troublesome relationship representations (e.g. a helpless 'infant' requiring total care from an adult, a self with exaggerated sense of power, and entitlement requiring constant confirmation from outside) that lie at the root of psychological disturbance. It is believed that the integration of these partial representations into more complex schemata, primarily but not exclusively through the use of insight, leads to improved internal and social adjustment.

Psychoanalysis is the most intensive form of these long-term therapies. The analysand attends treatment three or more times a week over a period of years. The use of the couch and the instruction to the analysand to free associate have been considered hallmarks. The distinction between psychoanalysis and other forms of psychotherapy is normally made in terms of the frequency of

sessions rather than in terms of the therapeutic stance of the analyst. It is difficult to avoid the conclusion that in the absence of plausible, theoretically based criteria for what is or is not psychoanalytic, against the background of an overwhelming diversity of theoretical frameworks, psychoanalysts have attempted to find common ground in readily identifiable treatment parameters. This problem arises as a consequence of an extremely loose relationship between psychoanalytic theory and clinical practice.[24] It is an indisputable fact that, whereas theory has evolved extremely rapidly in the last half of the twentieth century and continues to change, psychoanalytic practice has, until recently, changed surprisingly little and continues to provide the core of the psychoanalytic identity. On the other hand, the follow-along study by Sandell *et al.*[25] found that psychoanalysis and psychoanalytic psychotherapy were 'separate things'. When psychotherapy was performed using mainly psychoanalytic techniques, it was less effective than psychotherapy performed with modified and adjusted techniques (that is, not performed as an 'as-if analysis'). The findings from the Stockholm study suggest that psychoanalysis and psychoanalytic psychotherapy may be separate endeavours, although how exactly they differ is far from clear.

In this chapter we will not consider the theoretical richness of this field but instead will focus on the clinical constructs which run across the diverse intellectual approaches. The intersection of the two is perhaps clearest in one area which we shall consider in some detail—namely, the therapeutic action of long-term psychoanalytically oriented psychotherapeutic treatment.

Background

Historical development of the psychoanalytic approach to treatment

As is well known, Freud's discovery of the talking cure[26] was really that of an intelligent patient (Anna O) and her physician (Breuer). The patient reported that certain symptoms disappeared when she succeeded in linking up fragments of what she said and did in an altered state of consciousness (which we might now call dissociative) with forgotten impressions from her waking life. Breuer's remarkable contribution was that he had faith in the reality of the memories which emerged and did not dismiss the patient's associations as products of a deranged mind. The patient's response to treatment was probably less complete than Breuer and the young Freud had hoped[27] but the 'treatment' defined the basic elements of the 'cathartic' method-linking memory of trauma (the circumstances of her experience of her father's death) to her many symptoms.

At first Freud rigorously pursued the traumatogenic origins of neuroses. Later, when confronted by evidently incorrect statements, he modified his theory, assuming consistency between recollection and childhood psychic reality rather than physical reality.[28] The issue of accuracy of memories of childhood sexual trauma remains controversial, although its relevance to psychoanalytic technique is at best tangential.[29] Freud's technique, however, was dramatically modified by his discoveries. The intense emotional relationship between patient and physician, which had its roots in catharsis following hypnotic suggestion, had gradually subsided into what was principally an intellectual exercise to reconstruct the repressed causes of psychiatric disturbance from the fragments of material derived from the patient's associations. It was a highly mechanistic approach reminiscent of a complex crossword puzzle. In the light of therapeutic failures, however, Freud once more restored the emotional charge into the patient–physician relationship.[30] However, in place of hypnosis and suggestion, he used the patient's emotion, signs of transference of affect and affective resistance which were manifest in the analytic relationship. Instead of seeing the patient's intense emotional reaction to the therapist as an interference, Freud came to recognize the importance of transference as a representation of earlier relationship experiences which could make the reconstruction of those experiences in analysis highly meaningful to that individual.[31]

Freud's early clinical work evidently lacked some of the rigour which came to characterize classical psychoanalysis.[32] His occasional encouragement to his patients to join him on holiday might now be considered a boundary violation.[33] What is perhaps less well known is that Freud remained somewhat sceptical about the effectiveness of psychoanalysis as a method of treatment.[34] Indeed, autobiographies of some of his patients testify to his great flexibility as a clinician and use of non-psychoanalytic techniques, including behavioural methods.[35] Nor was Freud the only clinician to use psychoanalytic ideas flexibly. The Hungarian analyst Sandór Ferenczi should be credited with the discovery of the treatment of phobic disorders by relaxation and exposure[36] although many of his well-intentioned actions were criticized by contemporaries and more recently on arguable ethical grounds.[37]

The technique of psychoanalysis after Freud's death came to be codified. Those (such as Alexander and French and Freda Fromm-Reichmann) who attempted to revive or retain Freud's original clinical flexibility were subjected to powerful intellectual rebuttals.[38] In reality, psychoanalysts probably continued to vary in the extent to which they observed the ideals of therapeutic neutrality, abstinence, and a primarily interpretive stance, but these deviations could no longer be exposed to public scrutiny for fear of colleagues' forceful condemnation. Personal accounts of analyses with leading figures yield fascinating insights into variations in technique, principally in terms of the extent to which the analyst made use of a personal relationship.[39] There has been an ongoing dialectic throughout the history of psychodynamic approaches between those who emphasize interpretation and insight and those who stress the unique emotional relationship between patient and therapist as the primary vehicle of change. The controversy dates back to disputes concerning the work of Ferenczi and Rank[40] but re-emerged with the first papers of Balint and Winnicott in London opposing a Freudian and Kleinian tradition, and somewhat later in the United States with Kohut and more subtly Loewald opposing classical ego psychology.

In the last two decades, the pluralistic approach of modern psychoanalysis has brought out into the open many important dimensions along which psychoanalysts' techniques may vary. In particular, the recent trend to consider analyst and patient as equal partners engaged in a mutual exploration of meaning[41] directly challenged many of the classical constructs. The emphasis on the mutual influence of infant and caregiver shaped the emerging relational model of therapy as a two-person process in which there was little room for a detached analyst with pretensions of 'objectivity'. Drawing on the assumption that humans are predisposed towards two-person co-constructed systems that provide a context for psychic change, the quality of engagement

between therapist and patient became the core of therapeutic action. What changes the mind is not the insights gained but learning from the interactional experience of being with another person. Neither the analyst nor the patient can be considered as forging meaning; rather, meaning is co-constructed.

Technique—principal features

Neutrality and abstinence

Based in the classical framework of libidinal theory, Freud made an explicit injunction against the analyst giving in to the temptation of gratifying the patient's sexual desire.[42] Obviously, this is primarily an ethical issue. However, within the psychoanalytic context it also justifies the analyst's stance of resisting the patient's curiosity or using the therapeutic relationship in any way that consciously or unconsciously could be seen as motivated by the need to gratify their own hidden desires. Within this classical frame of reference, the patient must also agree to forgo significant life changes where these could be seen as relevant to current psychotherapeutic work. In practice, such abstinence on the part of the patient is rare. Yet long-term psychodynamic treatment may founder if the emotional experiences of the therapy are obscured by the upheavals of significant life events.

The primary function of abstinence is to ensure the neutrality of the therapist. The analyst assumes an attitude of open curiosity, empathy, and concern in relation to the patient. The therapist resists the temptation to direct the patient's associations and remains neutral irrespective of the subject matter of the patient's experiences or fantasies. While it is easy to take this issue too lightly, (and it is perhaps this aspect of the psychoanalyst's therapeutic stance which makes them most vulnerable to ridicule), it is probably genuinely critical for the therapist to retain emotional distance from the patient to a degree which enables the latter to bring fantasies and fears of which they feel uncertain. Nevertheless, neutrality at its worst denies the possibility of sensitivity; recent literature on the process and outcome of psychotherapy makes it clear that the therapist's genuine concern for the patient must become manifest if significant therapeutic change is to be achieved.[43] The quality of the alliance is one of the better predictors of outcome[44] and alliance is impacted by the patient's attachment style and quality of object-relations.[45]

Mechanisms of defence

The term 'psychic defences' may risk reification and anthropomorphism (precisely who is defending whom against what?) yet the existence of self-serving distortions of mental states relative to an external or internal reality is generally accepted, and frequently demonstrated experimentally.[46–48] Within classical psychoanalytical theory and its modern equivalent (ego psychology), intra-psychic conflict is seen as the core of mental functioning.[49] Here defences are seen as adaptations to reduce conflict. Within many object-relations theories, defences are seen as helpful to the individual to maintain an authentic or 'true' self-representation or a nuclear self.[17] Models of representations of relationships are of course often defensive. Traumatic experiences may give rise to omnipotent internal working models to address a feeling of helplessness. Within attachment theory, defences are construed as assisting in the maintenance of desirable relationships.[50] The Klein–Bion model makes limited use of the notion of defence mechanisms but uses the term in the context of more complex hypothetical structures called defensive organizations.[19] The term underscores the relative inflexibility of some defensive structures, which are thus best conceived of as personality types. For example, narcissistic personality disorder combines idealization and destructiveness; genuine love and truth are devalued. Such a personality type may have been protective to the individual at an earlier developmental stage, and has now acquired a stability or autonomy which must be rooted in the emotional gratification which such a self-limiting form of adaptation provides.[51]

Irrespective of the theoretical frame of reference, from a therapeutic viewpoint clinicians tend to differentiate between so-called primitive and mature defences based on the cognitive complexity entailed in their functioning.[52] In clinical work, primitive defences are often noted together in the same individual. For example, individuals loosely considered 'borderline' tend to idealize and then derogate the therapist. Thus they maintain their self-esteem by using splitting (clear separation of good from bad self-perception) and then projection. Projective identification[53] is an elaboration of the process of projection. An individual may ascribe an undesirable mental state to the other through projection but when the other can be unconsciously forced to accept the projection and experience its impact, the defence becomes far more powerful and stable. The analyst's experiencing of a fragment of the patient's self-state, has in recent years been considered an essential part of therapeutic understanding.[54]

Whether in fantasy or in actualized form, through projective identification the patient can experience a primitive mode of control over the therapist. Bion argued that when the self is experienced as being within another person (the therapist) the patient frequently attempts to exert total control over the recipient of the projection as part of an attempt to control split-off aspects of the self. Bion[55] also argued that not all such externalizations were of 'bad' parts of the self. Desirable aspects of the self may also be projected, and thus projective identification can be seen as a primitive mode of communication in infancy. There are other aspects of projective identification which we commonly encounter clinically. These include the acquisition of the object's attributes in fantasy, the protection of a valued aspect of the self from internal persecution through its evacuation into the object, and the avoidance or denial of separateness. It is thus a fundamental aspect of interpersonal relationship focused on unconscious fantasy and its appreciation is critical for the adequate practice of long-term psychotherapy.[56]

Classifications of defences have been frequently attempted[52,57–61] and often as a method for categorizing individuals or mental disorders.[4,5] An attachment theory-based classification rooted in the notion of habitual deactivation or hyperactivation of the attachment system ('attachment style') has achieved general acceptance.[62,63] Deactivating ('avoidant' or 'dismissing') strategies include suppression of ideas related to painful attachment experiences, repressing painful memories, minimizing stress and distress, segregated mental systems that result in the defensive exclusion of distressing material from the stream of consciousness.[64,65] Ingenious experimental studies have shown that individuals who habitually use avoidant defences are more efficient, when instructed, at suppressing conscious thoughts and associated feelings about a romantic partner leaving them for someone else[66] and are more likely to attribute their own unwanted traits to others

(projection) which serves to both increase self-other differentiation and enhance self-worth.[67] In a further, remarkable study the same group of researchers demonstrated that the above advantages of the suppression strategy of those using avoidant defence fall away in the laboratory situation if a cognitive load is placed on the participant which then leaves them literally defenceless so that they experience a heightened rebound of previously suppressed thought about painful separation.[68] The cognitive and socio-cognitive strategies associated with reducing anxiety or displeasure and enhancing safety, which both the attachment theory and psychoanalytic literatures tend to refer to as defences, are perhaps better thought of not as independent classes of mental activity or psychological entities but as a pervasive dynamic aspect of complex cognition interfacing with attachment relationships and emotional experience. Some mechanisms of defence are thought to be more characteristic of the less severe psychological disorders (e.g. depression, anxiety, obsessive–compulsive disorders, etc.). It is beyond the scope of this chapter to consider the various defence mechanisms in detail.

Modes of therapeutic action

The primary mode of the therapeutic action of psychoanalytic psychotherapy is generally considered to be insight.[69] Insight may be defined as the conscious recognition of the role of unconscious factors on current experience and behaviour. Unconscious factors encompass unconscious feelings, experiences, and fantasies. The psychodynamic model has been seen as a model of the mind that emphasizes repudiated wishes and ideas which have been warded off, defensively excluded from conscious experience. In our view this is a narrow and somewhat misleading way to define the therapeutic mechanism for approaches that are considered as psychodynamic. The psychodynamic approach is better seen as a stance taken to human subjectivity that is comprehensive, and aimed at understanding all aspects of the individual's relationship with her or his environment, external, and internal. Freud's great discovery ('where id was, there ego shall be', Freud[70] p. 80), often misinterpreted, points to the power of the conscious mind radically to alter its position with respect to aspects of its own functions, including the capacity to end its own existence through killing the body. Psychodynamic, in our view, refers to this extraordinary potential for dynamic self-alteration and self-correction—seemingly totally outside the reach of non-human species. Engaging with this potential to bring change through understanding, is the science and the art of the psychodynamic clinician.

Conscious insight is more than mere intellectual knowledge[71,72] or descriptive insights. Prototypically, psychodynamic therapy achieves demonstrated or ostensive insights which represent a more direct form of knowing, implying emotional contact with an event one has experienced previously. Working with what is non-conscious is at the heart of the dynamic approach to bringing about psychological change because of the force that awareness of unconscious expectations can bring to the interpretation of behaviour. Although specific formulations of the effect of insight depend on the theoretical framework in which explanations are couched, there is general agreement that insight has its therapeutic effect by in some way integrating mental structures.[72] Kleinian analysts[73] tend to see the healing of defensively created splits in the patient's representation of self and others as crucial. Split or part-objects may also be understood as isolated representations of intentional beings whose motivation is insufficiently well understood for these to be seen as coherent beings.[74] In this case insight could be seen as a development of the capacity to understand internal and external objects in mental state terms, thus lending them coherence and consistency.[75] The same phenomenon may be described as an increasing willingness on the part of the patient to see the interpersonal world from a third person's perspective.[76]

A simple demonstration to the patient of such an integrated picture of self or others is not thought to be sufficient.[31] The patient needs to 'work through' a newly arrived integration. Working through is a process of both unlearning and learning: actively discarding prior misconceptions and assimilating learning to work with new constructions. The technique of working through is not well described in the literature, yet it represents the critical advantage of long-term over short-term therapy.[77] Working through should be systematic and much of the advantage of long-term treatment may be lost if the therapist does not follow through insights in a relatively consistent and coherent manner.

In contrast to the emphasis on insight and working through are those clinicians who, as we have seen, emphasize the 'relationship aspect' of psychoanalytic therapy (Balint, Winnicott, Loewald, Mitchell, and many others). This aspect of psychoanalytic therapy was perhaps most eloquently described by Loewald when he wrote about the process of change as: 'set in motion, not simply by the technical skill of the analyst but by the fact that the analyst makes himself available for the development of a new 'object-relationship' between the patient and the analyst . . . '(Loewald, 1960, pp. 224–5).[78] Sandler and Dreher[79] have recently observed 'while insight is aimed for it is no longer regarded as an absolutely necessary requirement without which the analysis cannot proceed'. There is general agreement that the past polarization of interpretation and insight on the one hand, and bringing about change by presenting the patient with a new relationship on the other, was unhelpful. It seems that patients require both, and both may be required for either to be effective.[80]

Controversy remains even if all accept that neutrality is an impossible and undesirable fiction and that patient and therapist affect each other in myriad mutually influencing ways. Projective identification is seen as occurring in a bidirectional interpersonal field between analyst and patient—a model clearly adapted from Kleinian approaches to infant-caregiver interaction.[23] If we take this perspective seriously, we have to concede that all analytic interventions change the situations into which they are introduced, and their content and style always reflect the analyst's countertransference/response to the treatment situation.[81] Relational psychoanalysis advocates making the interactional influence of analyst upon patient explicit. As Levenson[82](p. 9) put it, the key therapeutic question is not 'what does this mean?' but rather 'what is going on around here?' The therapist will 'act' on the patient; this is not a therapeutic disaster but rather a potentially progressive and certainly inevitable part of the process.

It has been suggested that change in analysis will always be individualized according to the characteristics of the patient or the analyst.[83] For example, Blatt[84] suggested that patients who were 'introjective' (preoccupied with establishing and maintaining a viable self-concept rather than establishing intimacy) were more responsive to interpretation and insight. By contrast, anaclitic patients (more concerned with issues of relatedness than of self-development) were more likely to benefit from the quality of

the therapeutic relationship than from interpretation. Taking a second look at large-scale outcome investigations Blatt found strong evidence for the oft made but rarely demonstrated claim of patient personality—therapeutic technique fit.[85]

Indications and contraindications and selection procedures

Medical treatments normally have indications and contraindications. In psychodynamic treatment the term 'suitability' indicates a looser notion of the appropriateness of the approach.[86] Nevertheless, based primarily on clinical experience, some writers have arrived at specific criteria for long-term psychodynamic therapy.[87] Some authors have also suggested relatively systematic methods of assessment yielding both diagnostic and prognostic information.[88] The majority of psychodynamic clinicians, however, rely on clinical judgements based on interpersonal aspects of their first meeting with the patient.[71] The three areas of assessment are personal history, the content of the interview, and the style of the presentation.

A history of one good relationship has been traditionally regarded as a good indicator.[89] By contrast, a history of psychotic breakdown, severe obsessional states, somatization, and lack of frustration tolerance are generally considered contraindications. For example, a challenging set of re-analyses of the Treatment of Depression Collaborative Research Program found that the trait of perfectionism was associated with poor outcome, and could undermine the therapeutic alliance and the patient's satisfaction with social relations, limiting their improvement in the course of brief treatment for depression.[90]

Empirical literature, to the meagre extent that this is available, suggests that many of the presuppositions about suitability are unfounded. It was, for example, assumed that patients who manifested more serious mental illness, especially disturbances in reality testing, were unsuitable for psychoanalysis; however, a recent study showed that some patients with serious disturbances in reality testing were able to benefit from psychoanalysis when their analysts were able to tolerate and analyse this level of psychopathology.[91] What does seem to be consistent is that severity of symptoms, as well as functional levels in work and relationships, are correlated with the outcome of psychotherapy[92]—although no single patient variable is a strong predictor of outcome. This is why the effects of psychotherapy, good and bad, can sometimes be surprising.

Predicition based on the content of assessment interviews is hard. In general, the presence of some kind of 'mutuality' between therapist and patient is a positive indicator. Some clinicians offer 'trial interpretations' which summarize their initial impressions, and a positive thoughtful response to these is regarded a good indication. The capacity to respond emotionally within the assessment session is a further indicator.[93] Motivation for treatment is harder to ascertain. Most patients express enthusiasm for the treatment, which falls away once they are asked to confront unpleasant or unflattering parts of themselves.

More recently, psychodynamic therapists have given increasing consideration to the style of the patient's discourse during assessment rather than its content. Holmes,[94] for example, attempts to identify whether patients' narrative styles are avoidant (sparse and dismissing of interpersonal issues) or enmeshed and entangled (excessive current anger about past hurts and insults). The findings of one study indicate that, in a severely personality disordered population at least, the avoidant type of patient has a better prognosis in psychodynamic therapy.[95] A further relevant capacity is reflective function or mentalization, often reflected in narrative; this has been variously described as seeing oneself from the outside,[96] reflecting on one's inner world[87] or having fluidity of thought.[97]

Managing treatment

Starting treatment

(a) Establishing parameters

Most psychodynamic therapists, explicitly or implicitly, convey objectives and expectations to their patients. The details of this agreement normally include arrangements for a time and a place as well as the length and frequency of sessions. Usually a tentative idea is offered as to the likely duration of therapy: 'It is likely to take years rather than months.' Most therapists also describe the expected behaviour of the patient and the therapist: 'I would like you to be as open and honest with me as possible and say absolutely everything that comes into your mind. This is the fundamental rule.' In fact it is very likely, in view of the variety of such agreements that tend to be made, that its emotional context is more relevant than the specific items agreed upon. Such a 'contract' implies recognition by both patient and therapist that the process of therapy needs protecting and that it is important enough to require a sacrifice from both parties.

In the treatment of severe personality disorders, contracts may have an additional important function—that of protecting the therapy from incessant enactments, self-harming, parasuicidal gestures, and so on. In Kernberg's approach to the treatment of borderline patients, the patient formally undertakes not to seek the therapist's help outside of office hours, not to engage in acts of violence and to deal with self-destructive acts through normal medical channels.[98] Whilst such agreements are commonly made in long-term therapy, it is by no means clear that they are either essential or useful. For example, in an alternative form of psychodynamic therapy, Mentalization-Based Treatment (MBT), contracts are not recommended.[99]

(b) Formulation of patients' problems

An important part of initiating any psychosocial treatment is arriving at least at a preliminary formulation of the patient's problems. In the case of psychodynamic therapies this represents a special challenge because of the diversity of the possible theories to draw on. In principle, psychodynamic formulations would identify key unconscious conflicts, central maladaptive defences, unhelpful unconscious fantasies and expectations, deficits in personal development, and so on. The complexity of such formulations is such that agreements are hard to arrive at even when clinicians follow similar orientations. In the absence of a generally accepted format for formulating the patient's problems, a list of key parameters for the level of maturity of personality organization may be offered:

(a) the maturity of relationship representations (three or more persons versus just a self-other dimension)

(b) the maturity of psychic defences (primarily based on projective versus internalizing processes)

(c) the extent of whole as opposed to part object-relations (e.g. whether a person is represented as performing more than a single function for the patient)

(d) the general mutuality of the relationship patterns described; the quality of attachment to others.

It should be noted that psychodynamic formulations tend to change as treatment progresses. Indeed, Winnicott described psychoanalysis as 'an extended form of history taking'.[100] Within certain psychodynamic approaches formulation is communicated formally to patients (e.g. by letter in cognitive analytic therapy Ryle[101]).

The middle phase

(a) Supportive and directive interventions in psychodynamic therapy

Supportive techniques are used both explicitly and implicitly in psychodynamic treatment. They include offering explicit support and affirmation; offering reassurances concerning, for example, irrational anxieties about the therapeutic arrangements; expressing concern and sympathy to a patient who has suffered a recent loss; and general empathy for the patient's anxieties and struggles with the treatment.[102]

From a psychodynamic point of view, such supportive interventions are by no means straightforward. For example, Feldman[103] illustrated how patients may sometimes experience the therapist's submission to a demand for reassurance as a source of anxiety rather than comfort. They may be unconsciously aware that the therapist's true stance is not compatible with reassurance and therefore face anxieties about the therapist's weakness in allowing themselves to be manipulated. By contrast, Kohut's[17] emphasis on interpersonal empathy was probably a welcome antidote to the somewhat rigid interpretive stance of American ego psychologists, particularly for those whose history of psychosocial deprivation meant that they had experienced little by way of genuine warmth or concern in the past.

The most common use of supportive and directive techniques in psychodynamic psychotherapy are in the service of the therapy itself. Elaborative techniques (e.g. the simple question: 'Could you tell me more?') are undoubtedly directive in specifying a topic of interest, but at the same time may be crucial antecedents to interpretive work. Clarification stands in between supportive and interpretive interventions. It is a restatement in the therapist's words of the patient's communication. It may also be crucial in offering a verbal (symbolic) label for a confused set of internal experiences which the patient is poorly equipped to represent coherently. Confrontation is also in between a directive and an interpretive approach. At its gentlest, confrontation may involve the therapist simply identifying an inconsistency in the patient's communication and bringing this to the patient's attention. For example: 'You seem to express no sadness about this loss, yet in the past you claimed to have cared a great deal for him'.

(b) Regression

An important facet of psychoanalysis and long-term psychodynamic therapy is the activation and exploration of parts of the patient's personality which may be normally hidden behind an over-riding demand to adapt to the demands of every day life. Access to these aspects of personality is achieved through the process of regression. It has been suggested that rather than encouraging regression, the process is best conceived of as inhibiting 'an anti-regressive function' in much the same way that certain intimate interpersonal experiences, large group situations, and alcohol appear to bring out the more infantile aspects of our character.[104] Some psychoanalysts consider regression to be crucial to successful psychoanalytic treatment, but others consider the concept and its clinical application outmoded and counterproductive.[105] The extent to which a particular treatment involves significant regression appears to be a function of the patient's personality as well as the therapist's particular approach. Fear of regression is an important source of resistance to long-term psychotherapy, particularly amongst those with previous experience of psychotic episodes.[104]

(c) Resistance

Resistance is inevitably encountered in any long-term psychodynamic treatment. In fact, the presence of resistance is implied by the term dynamic, which suggests psychic forces both pulling against and pushing towards change. Like regression, resistance fluctuates in the middle stage of treatment. In borderline and narcissistic disorders, the patient's intense resistance signals the patient's desperation to protect extremely fragile self-esteem. In less severe cases, what appears to be at issue is preventing a painful integration of experience, such as the integration of love and hate directed towards the same object.[106]

In clinical practice resistance takes a variety of forms. In repression resistance, the patient may experience a temporary difficulty in gaining access to particular ideas and feelings; for example, failing to remember dreams. In transference resistance the patient may appear to wish to keep their relationship with their therapist at an extremely superficial level. In a negative therapeutic reaction the increase of symptomatology occurs alongside therapeutic progress. In Freud's formulation this may be attributed to unconscious guilt. It is quite likely that in at least some patients this form of resistance against psychotherapy is part of a pervasive so-called 'envious' predisposition to eradicate any aspect of their life that they experience as 'good' but beyond their immediate control.[107]

(d) The experience of the transference

Patients may experience a whole range of feelings about an analyst including love, admiration, excitement or anger, disappointment, and suspicion. The feelings appear to have little to do with the therapist's actual personality as different patients are likely to bring quite disparate feelings about the same analyst at the same time. While clearly not realistic, the actual nature of transference experience and its use in therapy is quite controversial.[108] Object-relations theorists consider the analyst a vehicle onto which an internal object (a person, an aspect of a person, the self, or an aspect of the self) is projected.[109] Clearly internal objects are representations which are heavily distorted by both fantasy and defensive processes.

For John Bowlby[64] transference feelings are based on expectations gathered through past relationship experience with an attachment figure. Patients resist understanding of the past relationship by insisting on repeating it. Bowlby's[110] suggestion that therapists function as secure bases implies that psychodynamic therapists are, in part, conducting attachment therapy as inevitably they serve as attachment figures for their patients. There is accumulating evidence for this claim[111–114] with a number of studies linking specific

transference schemas and attachment.[115–117] Many analysts do not accept such an isomorphism between past and present. Rather, they see it as something which gives coherence to the patient's experience of the analytic relationship—an aspect of narrative rather than a representation of the historical realities of the patient's experience.[118] In contrast, analysts who work in the Klein–Bion frame of reference see transference as providing an inevitably accurate picture of the patient's current internal world.[119] For example, a transference where the analyst is idealized may reflect psychotic anxieties in the patient linked to an intensification of the death instinct. The idealization serves to protect both the patient and the analyst from fantasized destruction which threatens to engulf them both. Marcia Cavell[120] demonstrated that these alternative models of transference have their philosophical roots in the debate between correspondence and coherence models of truth.

There is significant debate regarding from what point and how much psychoanalytic therapists should work 'in the transference'. Some analysts are inclined to see transference as pertinent to every aspect of the psychoanalytic situation. For example, Joseph[119] considers the therapeutic situation in toto as mirroring the internal state of the patient. Thus the therapeutic alliance or the 'real relationship'[121] are regarded as subsumed under the transference relationship. In this context it makes little sense to interpret anything other than the transference from the very beginning of the analysis. By contrast, Strachey[122] understood transference as an attempted externalization of the patient's superego. Unlike other people in the patient's life, the analyst does not accept this externalization, whether it is idealized, denigratory, or judgemental. The analyst conveys his or her understanding of the externalization by a so-called 'mutative interpretation'. While Strachey implied that only interpretation of the transference is therapeutic, his view clearly admits other aspects of the therapeutic relationship. Other therapists, particularly Freudian psychoanalysts, regard transference interpretations as an important but not uniquely therapeutic way of providing the patient with insight and consider the almost exclusive reliance on understanding the patient through their thoughts and feelings about their therapists as unhelpful and even dangerous.[123] The only systematic investigation of this technical controversy, where patients were randomly assigned to a transference and a non-transference-oriented psychological therapy, could not show a significant difference between the overall effectiveness of these two treatments, although there was a tendency for those with more dysfunctional object-relationship representations to do better in therapy which used transference interpretations.[124,125]

The nature of the transference appears to systematically relate to specific clinical groups and hence may have an aetiological significance. For example, specific transference patterns appear to characterize particular groups of narcissistic patients.[17] The 'mirroring' transference is one where patients crave the approbation and admiration of the therapist. This may be a consequence of the failure of the original self-objects (parents) in their mirroring function. If this transference is undermined by premature interpretations, an opportunity for restoring self-esteem is lost. The 'idealizing' transference also enables the patient to address a deficiency in self-esteem by secretly identifying with the object of admiration (the analyst). If the analyst destroys this idealized image, within Kohut's framework, this is equivalent to a direct attack on the patient's self-regard. Other analysts would suspect that behind such an exaggeratedly positive image lies the patient's true image of the analyst as frustrating or inadequate, an image which is simply placed out of harm's way by the idealization. An interesting empirical study of clinicians' experience of the transference with personality disordered patients was reported from Drew Westen's laboratory.[115] The study identified five transference dimensions: angry/entitled, anxious/preoccupied, avoidant/counterdependent, secure/engaged, and sexualized which were associated in predictable ways with Axis II pathology and confirmed that the way patients interact with their therapists can provide important data about their personality, attachment patterns, and interpersonal functioning.

Commonly, transference includes an erotic component, regardless of the age or even the gender of the analyst.[126] Admitting to such feelings may border on the unacceptable for some patients. Attachment theorists may suggest that sexual fantasies are used in the service of obtaining the attention of an unresponsive attachment figure.[127] Eroticized transference, relatively common in severely traumatized patients, represents an expression of a need for sexual gratification which, in the context of the therapy, is not considered by the patient as unrealistic.[71] Some view this phenomenon as an indication of an immature mode of representing internal reality, where only the physically observable outcome is believed to be real.[128]

(e) Experience of the countertransference

Countertransference is a somewhat controversial concept in psychoanalytic clinical work. The therapist during the course of an intensive long-term treatment is likely to have a range of feelings which are related to the patient's current experience but which may serve to either illuminate or obscure this. Some countertransference experiences may be instances of projective identification and thus can be appropriately attributed to the patient,[129] whereas others are likely to be the analyst's neurotic emotional reactions to the patient's behaviour or the material he or she brings. For Freud,[130] countertransference was always of this latter type, a neurotic reaction which was likely to obstruct psychoanalytic treatment. It was not until Paula Heimann[131] pointed out that the analyst's feelings and thoughts could contain important clues about the patient's unconscious mental state that countertransference started to be seriously considered as part of the analyst's therapeutic armamentarium. Those following an interpersonalist tradition saw the recognition of the complementarity of the therapeutic relationship as highly appropriate. From this point of view, the assumption of perfect neutrality on the part of the analyst who is a participant as well as an observer is both an anathema and an anachronism.[132] The psychotherapeutic process is more accurately viewed as a complex mixture of complementary interpersonal processes which establish themselves in 'custom designed' configurations in each treatment.[133]

The therapist's feelings may be either complementary to or concordant with those of the patient.[134] Concordant countertransferences are the product of primitive, empathic processes within the therapist who 'feels' for the patient, who may unconsciously react to experiences implied but not yet verbalized by the patient; for example, inexplicable overwhelming sadness. Complementary countertransferences tend to occur when the patient treats the analyst in a manner consistent with interpersonal interactions within a past relationship. Most commonly this occurs when the patient treats the therapist as he or she experienced being treated as a child. This is known as the 'reverse transference'.[135]

The mechanisms of countertransference are poorly understood. To assert that countertransference functions via projective identification merely brings one poorly understood phenomenon to account for a second even less well understood one. Sandler[136] suggested that an instantaneous process of automatic mirroring of one's partner in an act of communication accounted for concordant countertransference. The process, which he termed primary identification, was non-conscious and could be brought into awareness only upon reflection. Recent work on the mirror neurone system[137,138] suggests that the fundamental mechanism that allows us to understand the actions and emotions of others involves the activation of the mirror neurone system for actions and the activation of visceromotor centres for the understanding of affect. An alternative account suggests that a secondary mode of encoding is available within language whereby the use of a language of pretend gestures at the phonemic, syntactic, or even semantic level enables the communicator to address directly the unconscious of the recipient of the communication.[139] In other words, anything that can be said in gestures may be communicated unconsciously through language, through phonemic distortion, intonation, and other paralinguistic features and picked up impressionistically by the therapist.

When either concordant or complementary countertransferences mobilize defensive processes within the analyst, countertransference is in danger of becoming disruptive to therapeutic understanding. The analyst may react by unconsciously withdrawing from the therapeutic relationship. For example, in the case of a concordant countertransference where the patient's feelings of inadequacy create a similar feeling in the analyst, the analyst's vulnerability in this area may lead him or her to become defensively angry or excessively motivated to demonstrate his or her efficacy. There may be no simple way of regulating such reactions and the only reasonable strategy might be to carefully monitor one's style of relating, noting anything that is unusual. A number of analysts have pointed to the importance of reflectiveness in this context.

Some feelings in relation to the patient are not provoked either by the patient's projections or the neurotic feelings these give rise to in the therapist. It required someone of the stature of Donald Winnicott[140] to make the self-evident observation that the provocative behaviour of certain patients (particularly those in the borderline spectrum) can lead to a normal reaction of 'objective hate'. These reactions are merely indications of the therapist's humanity. Analytic understanding of these sometimes intense reactions to patients helps, but models of countertransference ill-fit such experiences. The objective study of countertransference has had to wait for a recent ingenious methodological development from Westen's laboratory.[141] The Countertransference Questionnaire yielded eight clinically and conceptually coherent factors that were independent of clinicians' theoretical orientation: (i) overwhelmed/disorganized, (ii) helpless/inadequate, (iii) positive, (iv) special/overinvolved, (v) sexualized, (vi) disengaged, (vii) parental/protective, and (viii) criticized/mistreated. Countertransference patterns were systematically related to patients' personality pathology across therapeutic approaches, suggesting that clinicians, regardless of therapeutic orientation, can make diagnostic and therapeutic use of their own responses to the patient.

(f) Interpretation

Interpretive interventions are at the core of psychoanalytic and psychodynamic treatment. However, the importance of interpretation is often exaggerated in relation to other aspects of the therapy. It is a sobering reminder that follow-up studies of long-term psychodynamic therapies invariably demonstrate that patients remember their analyst not for their interpretive interventions, rarely remembering individual interpretations, but rather for their 'emotional presence', regardless of the analyst's therapeutic perspective.[142]

Interpretations may be classified according to the aspect of a conflict they aim to address: the defence, the anxiety, or the underlying wish or feeling. Similarly, the content of the interpretation may be used in classifying interpretations: whether it relates to external reality, the transference relationship, or childhood relationships. In principle, in the earliest phases of treatment interpretations relating to current events are most common and, as the treatment progresses, transference issues and the patient's past may increasingly take over as foci of analytic work. Interpretations should start with the patient's anxiety, by identifying the defence used by the patient to protect himself from repudiated wishes and affects. In reality, these are guidelines that are rarely followed in practice. For example, very long-term treatments tend to end up being principally supportive explorations of the patient's current experience.[143] Furthermore, interpretations of the distant past tend to be least helpful to individuals with severe personality disorders.[144] Working in the so-called 'here and now' is more effective with those patients whose representation of the past is unreliable and distorted.[145]

Steiner[146] distinguished analyst-centred from patient-centred interpretations. The former refers to comments on the patient's reactions in terms of what the patient thinks may be going on in the analyst's mind, while the latter directly addresses the analyst's perception of the patient's non-conscious mental state. In either case the patient is directly learning about how minds interact in the context of social relationships. The distinction is important since when patient-centred interpretations are used exclusively the therapist may appear to be persecutory and not to be cognizant of the patient's genuine difficulties in being in an intimate relationship with another person. Others have argued, that at least in the case of severe personality disorder, interpretations, if they were to have therapeutic value, should focus on the patient's understanding of thoughts and feelings in themselves or in others at the level of what was conscious rather than unconscious, what patients could discover for themselves rather what they received as a communication from a 'mind expert'.[147] This implies that interpretation of the transference is about helping the patient represent their own and their therapist's mental states in the treatment room in all their complexity but with a stance conveying enquiry and playful curiosity about something that is not readily knowable (the mental state of the other is always opaque) with the aim of making thinking about thoughts and feelings safe again rather than communicating powerful insights.

The idealization of the transference has led some therapists to neglect interpretation of the patient's behaviour outside of the therapy. Most clinicians now agree that a balance needs to be struck between these two approaches. Treatment which is over-focused on the transference becomes a claustrophobic enclave.[148] In certain instances, the direct communication of the therapist's experience of frustration (objective hate in Winnicott's terms) may help to break a rigid repetitive pattern in the therapy.[149] Disclosing the therapist's experience is one of the cutting edges of the relational

approach to psychodynamic therapy.[150] In cases where the therapeutic alliance falters, perhaps following an empathic failure on the part of the therapist, it turns out that the recovery of the alliance may have particular therapeutic value both in showing the possibility of repair[151] but also as an opportunity to understand misunderstanding, an ideal opportunity for the recovery of mentalization.[152]

Ending treatment

The ending of psychoanalytic therapy is often idealized in clinical descriptions. As there is little agreement on the goals of psychoanalytic therapy,[79] it is hardly surprising that there is little general agreement about when ending is appropriate. Desirable final outcomes are mostly stated in terms of the process of treatment and are thus mostly specified in theoretical terms (e.g. increased awareness of impulses and fantasies, a reintegration of aspects of the self lost through projective identification, the capacity to engage in self-analysis, etc.). All these, even if observable in the course of treatment, are only loosely related to the aims the patient might have in concluding a lengthy treatment process.

The patient's own goals tend to be outcome rather than process goals and are more easily defined: the decline of symptoms, improved relationships, greater well-being, increased capacity for work, higher self-esteem, a capacity for assertiveness. As such changes are clearly achievable without psychodynamic treatment, many psychodynamic clinicians erroneously regard such criteria for ending as superficial. Independent evidence will be required to show that the achievement of process aims results in a more permanent or general achievement of outcome aims, in order to validate process aims as an appropriate criterion for ending.

Ending itself, of course, is a process. There is significant disagreement between authors, however, as to its nature; it has been labelled among other things as a mourning,[153] a detachment,[71] and a maturation.[154] It is inevitable that there is disappointment and disillusionment at the ending of long-term therapy as what is achieved is never quite the same as what has been hoped for.[155] Also, the patient loses the object who has been available as a receptacle for projections.[146] It is not surprising then, that symptoms sometimes return, even if only briefly, as part of the process of termination and the full benefit is not seen until some months after termination.[25] There is general agreement, however, that with these unconscious issues worked through the ending of therapy requires no special form of intervention on the part of the therapist.

Efficacy

It is often said that there are no studies on the effectiveness of psychoanalysis and long-term psychodynamic psychotherapy. In fact, this is not true. There are a number of comprehensive reviews[156–160] and they tend to come to similar conclusions. There is considerable evidence for the effectiveness of psychoanalytic approaches but definitive randomized controlled trials of its efficacy are still lacking.

The Boston Psychotherapy study[161] compared long-term psychoanalytic therapy (two or more times a week) with supportive therapy for clients with schizophrenia in a randomized controlled design. On the whole clients who received psychoanalytic therapy fared no better than those who received supportive treatment. In a partial-hospital RCT[162,163] the psychoanalytic arm of the treatment included therapy groups three times a week as well as individual therapy once or twice a week over an 18 month period.

The Stockholm Outcome of Psychotherapy and Psychoanalysis Project[164–166] followed 756 persons who received national insurance funded treatment for up to 3 years in psychoanalysis or psychoanalytic psychotherapy. The groups were matched on many clinical variables. Four or five times weekly analysis had similar outcomes at termination when compared with one to two sessions per week psychotherapy. During the follow-up period, psychotherapy patients did not change but those who had had psychoanalysis continued to improve, almost to a point where their scores were indistinguishable from those obtained from a non-clinical Swedish sample.

The German Psychoanalytic Association undertook a major follow-up study ($n = 401$) of psychoanalytic treatments undertaken in that country between 1990 and 1993.[159,167] Between 70 per cent and 80 per cent of the patients achieved (average 6.5 years after the end of treatment) good and stable psychic changes according to the evaluations of the patients, their analysts, independent psychoanalytic and non-psychoanalytic experts, and questionnaires commonly applied in psychotherapy research. The evaluation of mental health costs showed a cost reduction through fewer days of sick leave during the 7 years following the end of long-term psychoanalytic treatments. In the absence of pre-treatment measures it is impossible to estimate the size of the treatment effect.

The Research Committee of the International Psychoanalytic Association recently prepared a comprehensive review of North American and European outcome studies of psychoanalytic treatment.[157] Four case record studies, 13 naturalistic pre-post or quasi-experimental studies, nine follow-up studies, and nine experimental studies were identified. In addition, six process-outcome studies were also reviewed. The committee concluded that existing studies failed to demonstrate unequivocally the efficacy of psychoanalysis relative to either alternative treatment or active placebo. Studies showed a range of methodological and design problems including absence of intent to treat controls, heterogeneous patient groups, lack of random assignments, failure to use independently administered standardized measures of outcome, etc.

Another overview[168] suggested that psychoanalytic treatments may be necessary when other treatments proved to be ineffective. The authors concluded that psychoanalysis appears to be consistently helpful to patients with milder disorders and somewhat helpful to those with more severe disturbances. More controlled studies are necessary to confirm these impressions. A number of studies testing psychoanalysis with 'state of the art' methodology are ongoing and are likely to produce more compelling evidence over the next years. Despite the limitations of the completed studies, evidence across a significant number of pre-post investigations suggests that psychoanalysis appears to be consistently helpful to patients with milder (neurotic) disorders and somewhat less consistently so for other, more severe groups. Across a range of uncontrolled or poorly controlled cohort studies, mostly carried out in Europe, longer intensive treatments tended to have better outcomes than shorter, non-intensive treatments (demonstration of a dose-effect relationship). The impact of psychoanalysis was apparent beyond symptomatology, in measures of work functioning and reductions in health care costs. Studies report results which other psychotherapies have not been able to achieve; some studies show very long-term benefits from psychoanalytic treatment; the

results tend to be highly consistent across studies; some of the populations studied have been larger than most better controlled treatment trials. So whereas it is true to say that little that is definite can be stated about the outcome of psychoanalysis, a number of suggestive conclusions may be drawn and these are listed below.

Across a number of studies and measures psychoanalysis has been shown to benefit the majority of those who are offered this treatment[169] and can bring the functioning of a clinical group to the level of the normal population.[167] Completed treatments tend to be associated with greater benefits.[170] On the whole longer treatments have better outcomes[171] and intensive psychoanalytic treatment is generally more effective than psychoanalytic psychotherapy,[25] but its superiority sometimes only becomes apparent on long-term follow-up.[172] Psychoanalysis can lead to a reduction in health care related use and expenditure[173] and this is maintained for a number of years after therapy ends[174] but it does not invariably achieve this.[166] Psychoanalytic treatment can lead to a reduction in the use of psychotropic medication amongst inpatients.[175] Long-term psychoanalytic therapy can reduce symptomatology in severe personality disorders such as BPD[162,176,177]and these improvements are maintained.[163]

Training

Training in psychoanalytic psychotherapy and psychoanalysis has three components: a personal psychoanalytic psychotherapy, theoretical training, and supervised clinical practice. A variety of trainings are available, although in most countries there is only one training organization that is recognized by the International Psychoanalytic Association. Training is long, chiefly because of the length of supervised treatments. Training standards are carefully monitored by national and international bodies.

Conclusion

Psychoanalysis is hardly a practical treatment alternative for the twenty-first century. The principles derived from this treatment, however, have powerfully influenced other psychotherapeutic approaches, whether long-term or short-term therapy or psychiatric care more generally, particularly in the United States. At the time of its invention, it was the unique effective psychosocial treatment method for psychiatric disorder which offered a genuine alternative to the sometimes barbaric and generally ineffective treatment methods available. Not surprisingly, its proponents adopted an almost religious zeal in defending its value against alternative approaches. While understandable, such an attitude has no place in the sophisticated evidence base underpinning multi-agency service planning. Psychoanalytic clinicians face a challenge in identifying their niche in the complex mental health care delivery systems of the twenty-first century.

Further information

Budd, S. and Rusbridger, R. (eds.) (2005). *Introducing psychoanalysis: essential themes and topics*. Routledge, London.

Fonagy, P. and Target, M. (2003). *Psychoanalytic theories: perspectives from developmental psychopathology*. Whurr, London.

Yeomans, F.E., Clarkin, J.F., and Kernberg, O.F. (eds.) (2002). *A primer of transference-focused psychotherapy for the borderline patient*. Jason Aronson, Northvale, NJ.

References

1. Fonagy, P. and Target, M. (2003). *Psychoanalytic theories: perspectives from developmental psychopathology*. Whurr, London.
2. Smith, H.F. (2003b). Conceptions of conflict in psychoanalytic theory and practice. *The Psychoanalytic Quarterly*, **72**, 49–96.
3. Sandler, J. (2003). On attachment to internal objects. *Psychoanalytic Inquiry*, **23**, 12–26.
4. Bond, M. (2004). Empirical studies of defense style: relationships with psychopathology and change. *Harvard Review of Psychiatry*, **12**(5), 263–78.
5. Lenzenweger, M.F., Clarkin, J.F., Kernberg, O.F., *et al.* (2001). The inventory of personality organization: psychometric properties, factorial composition, and criterion relations with affect, aggressive dyscontrol, psychosis proneness, and self-domains in a nonclinical sample. *Psychological Assessment*, **13**(4), 577–91.
6. Fonagy, P., Target, M., and Gergely, G. (2006). Psychoanalytic perspectives on developmental psychopathology. In *Developmental psychopathology: theory and methods*, Vol. 1 (2nd edn) (eds. D. Cicchetti and D.J. Cohen), pp. 701–49. John Wiley & Sons, Inc., New York.
7. Brumbaugh, C.C. and Fraley, R.C. (2006). Transference and attachment: how do attachment patterns get carried forward from one relationship to the next? *Personality and Social Psychology Bulletin*, **32**(4), 552–60.
8. Adler, G. and Buie, D. (1979). Aloneness and borderline psychopathology: the possible relevance of some child developmental issues. *The International Journal of Psycho-analysis*, **60**, 83–96.
9. Eagle, M. (2003). Clinical implications of attachment theory. *Psychoanalytic Inquiry*, **23**(1), 27–53.
10. Mikulincer, M. and Shaver, P.R. (2004). Security-based self representations in adulthood: contents and processes. In *Adult attachment: theory, research and clinical implications* (eds. W.S. Rholes and J.A. Simpson), pp. 159–95. Guilford, New York.
11. Winnicott, D.W. (1958). The capacity to be alone. In *The maturational processes and the facilitating environment*, pp. 29–36. International Universities Press, New York, 1965.
12. Westen, D. and Gabbard, G.O. (2002). Developments in cognitive neuroscience. II. Implications for theories of transference. *Journal of the American Psychoanalytic Association*, **50**(1), 99–134.
13. Hartmann, H. (1939). *Ego psychology and the problem of adaptation*. International Universities Press, New York, 1958.
14. Freud, A. (1965). *Normality and pathology in childhood: assessments of development*. International Universities Press, Madison, CT.
15. Klein, M. (1948). On the theory of anxiety and guilt. In *Envy and gratitude and other works, 1946–1963*. (eds. M. Masud and R. Khan) Delacorte Press, New York, 1975.
16. Fairbairn, W.R.D. (1952). *An object-relations theory of the personality*. Basic Books, New York, 1954.
17. Kohut, H. (1984). *How does analysis cure*? University of Chicago Press, Chicago.
18. Kernberg, O.F. (1976). *Object relations theory and clinical psychoanalysis*. Aronson, New York.
19. Rosenfeld, H. (1987). *Impasse and interpretation*. Tavistock Publications, London.
20. Aron, L. and Harris, A. (eds.) (2005). *Relational psychoanalysis: innovation and expansion*, Vol. II. Analytic Press, Hillsdale, NJ.
21. Sullivan, H.S. (1953). *The interpersonal theory of psychiatry*. Norton, New York.
22. Mitchell, S.A. and Aron, L. (eds.) (1999). *Relational psychoanalysis: the emergence of a tradition*. Analytic Press, Hillsdale, NJ.
23. Seligman, S. (2003). The developmental perspective in relational psychoanalysis. *Contemporary Psychoanalysis*, **39**, 477–508.
24. Fonagy, P. (2003). Some complexities in the relationship of psychoanalytic theory to technique. *The Psychoanalytic Quarterly*, **72**, 13–48.

25. Sandell, R., Blomberg, J., and Lazar, A. (2002). Time matters. On temporal interactions in psychoanalysis and long-term psychotherapy. *Psychotherapy Research*, **12**, 39–58.
26. Freud, S. and Breuer, J. (1895). Studies on hysteria. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 2 (ed. J. Strachey), pp. 1–305. Hogarth Press, London.
27. Castelnuovo-Tedesco, P. (1994). On rereading the case of Anna O: more about questions that are unanswerable. *Journal of the American Academy of Psychoanalysis*, **22**, 57–71.
28. Freud, S. (1899). Screen memories. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 3 (ed. J. Strachey), pp. 301–22. Hogarth Press, London.
29. Fonagy, P. and Target, M. (1997). Perspectives on the recovered memories debate. In *Recovered memories of abuse: true or false*? (eds. J. Sandler and P. Fonagy), pp. 183–216. Karnac Books, London.
30. Freud, S. (1912a). The dynamics of transference. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 12, (ed. J. Strachey) pp. 97–109. Hogarth Press and the Institute of Psycho-Analysis, London.
31. Freud, S. (1914). Remembering, repeating, and working through. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 12 (ed. J. Strachey), pp. 145–56. Hogarth Press, London.
32. Jones, E. (1953). *The life and work of Sigmund Freud*, Vol. I. Basic Books, New York.
33. Celenza, A. and Gabbard, G.O. (2003). Analysts who commit sexual boundary violations: a lost cause? *Journal of the American Psychoanalytic Association*, **51**(2), 617–36.
34. Freud, S. (1937). Analysis terminable and interminable. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 23 (ed. J. Strachey), pp. 209–53. Hogarth Press, London.
35. Walter, B. (1946). *Theme and variations*. Knopf, New York.
36. Ferenczi, S. (1930). The principle of relaxation and neocatharsis. *The International Journal of Psycho-analysis*, **11**, 428–43.
37. Szecsödi, I. (2007). Sándor Ferenczi–the first intersubjectivist. *Scandinavian Psychoanalytic Review*, **30**(1), 31–41.
38. Eissler, K.R. (1953). The effect of the structure of the ego on psychoanalytic technique. *Journal of the American Psychoanalytic Association*, **1**, 104–43.
39. Guntrip, H. (1975). My experience of analysis with Fairbairn and Winnicott. *International Review of Psychoanalysis*, **2**, 145–56.
40. Ferenczi, S. and Rank, O. (1925). *The development of psychoanalysis*. International Universities Press, Madison, CT, 1986.
41. Altman, N., Briggs, R., Frankel, J., *et al.* (2002). *Relational child psychotherapy*. The Other Press, New York.
42. Freud, S. (1915). Observations on transference love. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 12 (ed. J. Strachey), pp. 157–71. Hogarth Press, London.
43. Lambert, M. (ed.) (2004). *Bergin and Garfield's handbook of psychotherapy and behavior change.* Wiley, New York.
44. Orlinksy, D.E., Ronnestad, M.H., and Willutski, U. (2004). Fifty years of psychotherapy process-outcome research: continuity and change. In *Bergin and Garfield's handbook of psychotherapy and behavior change* (ed. M. Lambert), pp. 307–90. Wiley, New York.
45. Pinsker-Aspen, J., Stein, M., and Hilsenroth, M. (2007). Clinical utility of early memories as a predictor of early therapeutic alliance. *Psychotherapy: Theory, Research, Practice, Training*, **44**, 96–109.
46. Blagov, P.S. and Singer, J.A. (2004). Four dimensions of self-defining memories (specificity, meaning, content, and affect) and their relationships to self-restraint, distress, and repressive defensiveness. *Journal of Personality*, **72**(3), 481–511.
47. Lyons-Ruth, K. (2003). Dissociation and the parent-infant dialogue: a longitudinal perspective from attachment research. *Journal of the American Psychoanalytic Association*, **51**(3), 883–911.
48. Shamir-Essakow, G., Ungerer, J.A., Rapee, R.M., *et al.* (2004). Caregiving representations of mothers of behaviorally inhibited and uninhibited preschool children. *Developmental Psychology*, **40**(6), 899–910.
49. Brenner, C. (1982). *The mind in conflict*. International Universities Press, New York.
50. Walins, D.J. (2007). *Attachment and psychotherapy*. Guilford Press, New York.
51. Steiner, J. (2000). Containment, enactment and communication. *The International Journal of Psycho-analysis*, **81**(2), 245–55.
52. Vaillant, G.E. (1992). *Ego mechanisms of defense: a guide for clinicians and researchers*. American Psychiatric Association Press, Washington, DC.
53. Klein, M. (1946). Notes on some schizoid mechanisms. In *Developments in psychoanalysis* (eds. M. Klein, P. Heimann, S. Isaacs, and J. Riviere), pp. 292–320. Hogarth Press, London.
54. Heimann, P. (1956). Dynamics of transference interpretation. *International Journal Psycho-analysis*, **37**, 303–10.
55. Bion, W.R. (1962). *Learning from experience*. Heinemann, London.
56. Greatrex, T. S. (2002). Projective identification: how does it work? *Neuro-Psychoanalysis*, **4**, 187–97.
57. Fraiberg, S. (1982). Pathological defenses in infancy. *The Psychoanalytic Quarterly*, **51**, 612–35.
58. Freud, A. (1936). *The ego and the mechanisms of defence*. International Universities Press, New York, 1946.
59. Horowitz, M.J. (1995). Defensive control states and person schemas. In *Research in psychoanalysis: process, development, outcome* (eds. T. Shapiro and R.N. Emde), pp. 67–89. International Universities Press, Madison, CT.
60. Kaye, A.L. and Shea, M.T. (2000). Personality disorders, personality traits, and defense mechanisms. In *Handbook of psychiatric measures* (ed. Task Force for the Handbook of Psychiatric Measures), pp. 713–49. American Psychiatric Association, Washington, DC.
61. Spitz, R. (1961). Some early prototypes of ego defenses. *Journal of the American Psychoanalytic Association*, **9**, 626–51.
62. Cassidy, J. and Kobak, R.R. (1988). Avoidance and its relation to other defensive processes. In *Clinical implications of attachment* (eds. J. Belsky and T. Nezworski), pp. 300–23. Erlbaum, Hillsdale, NJ.
63. Mikulincer, M. and Shaver, P.R. (2003). The attachment behavior system in adulthood: contents and processes. In *Advances in experimental social psychology*, Vol. 35 (ed. M.P. Zanna), pp. 53–152. Academic Press, San Diego, CA.
64. Bowlby, J. (1980). *Attachment and loss, Vol. 3: loss: sadness and depression*. Hogarth Press and Institute of Psycho-Analysis, London.
65. George, C. and West, M. (2001). The development and preliminary validation of a new measure of adult attachment: the adult attachment projective. *Attachment & Human Development*, **3**(1), 30–61.
66. Fraley, R.C. and Shaver, P.R. (1997). Adult attachment and the suppression of unwanted thoughts. *Journal of Personality and Social Psychology*, **73**(5), 1080–91.
67. Mikulincer, M. and Horesh, N. (1999). Adult attachment style and the perception of others: the role of projective mechanisms. *Journal of Personality and Social Psychology*, **76**(6), 1022–34.
68. Mikulincer, M., Dolev, T., and Shaver, P.R. (2004). Attachment-related strategies during thought suppression: ironic rebounds and vulnerable self-representations. *Journal of Personality and Social Psychology*, **87**(6), 940–56.
69. PDM Task Force. (2006). *Psychodynamic diagnostic manual*. Alliance of Psychoanalytic Organizations, Silver Spring, MD.
70. Freud, S. (1933). New introductory lectures on psychoanalysis. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 22 (ed. J. Strachey), pp. 1–182. Hogarth Press, London.
71. Etchegoyen, H. (1991). *The fundamentals of psychoanalytic technique*. Karnac, London.
72. Thomä, H. and Kächele, H. (1987). *Psychoanalytic practice. I. Principles*. Springer-Verlag, New York.

73. Spillius, E.B. (2001). Freud and Klein on the concept of phantasy. *The International Journal of Psycho-analysis*, **82**(2), 361–73.
74. Gergely, G. (2000). Reapproaching Mahler: new perspectives on normal autism, normal symbiosis, splitting and libidinal object constancy from cognitive developmental theory. *Journal of the American Psychoanalytic Association*, **48**(4), 1197–228.
75. Allen, J.G. (2006). Mentalizing in practice. In *Handbook of mentalization based treatments* (eds. J.G. Allen and P. Fonagy), pp. 3–30. Wiley, Chichester.
76. Britton, R. (1998). *Belief and imagination*. Routledge, London.
77. Lipsius, S.H. (2001). Working through in psychoanalytic psychotherapy: an alternative and complementary path. *Journal of the American Academy of Psychoanalysis*, **29**, 585–600.
78. Loewald, H.W. (1960). On the therapeutic action of psycho-analysis. *The International Journal of Psycho-analysis*, **41**, 16–33.
79. Sandler, J. and Dreher, A.U. (1996). *What do psychoanalysts want? The problem of aims in psychoanalysis*, Vol. 24. Routledge, London and New York.
80. Chodorow, N.J. (2003). The psychoanalytic vision of Hans Loewald. *The International Journal of Psycho-analysis*, **84**, 897–913.
81. Hoffman, I.Z. (2006). The myths of free association and the potentials of the analytic relationship. *The International Journal of Psycho-analysis*, **87**(Pt 1), 43–61.
82. Levenson, E. (1983). *The ambiguity of change*. Basic Books, New York.
83. Pine, F. (1998). *Diversity and direction in psychoanalytic technique*. Yale University Press, New Haven, CT.
84. Blatt, S.J. (2004). *Experiences of depression: theoretical, clinical and research perspectives*. American Psychological Association, Washington, DC.
85. Blatt, S.J., Auerbach, J.S., Zuroff, D.C., *et al.* (2006). Evaluating efficacy, effectiveness, and mutative factors in psychodynamic psychotherapies. In *Psychodynamic diagnostic manual* (ed. PDM Task Force). Alliance of Psychoanalytic Organizations, Silver Spring, MD.
86. Varvin, S. (2003). Which patients should avoid psychoanalysis, and which professionals should avoid psychoanalytic training? A critical evaluation. *Scandinavian Psychoanalytic Review*, **26**, 109–22.
87. Coltart, N. (1988). Diagnosis and assessment for suitability for psycho-analytic psychotherapy. *British Journal of Psychotherapy*, **4**, 127–134.
88. Kernberg, O.F. (1981). Structural interviewing. *The Psychiatric Clinics of North America*, **4**, 169–95.
89. Piper, W.E., Ogrodniczuk, J.S., McCallum, M., *et al.* (2003). Expression of affect as a mediator of the relationship between quality of object relations and group therapy outcome for patients with complicated grief. *Journal of Consulting and Clinical Psychology*, **71**(4), 664–71.
90. Shahar, G., Blatt, S.J., Zuroff, D.C., *et al.* (2003). Role of perfectionism and personality disorder features in response to brief treatment for depression. *Journal of Consulting and Clinical Psychology*, **71**(3), 629–33.
91. Leuzinger-Bohleber, M. (2002). A follow-up study critical inspiration for our clinical practice? In *Outcomes of psychoanalytic treatment. Perspectives for therapists and researchers* (eds. M. Leuzinger-Bohleber and M. Target). Whurr Publishers, London and Philadelphia.
92. Clarkin, J.F. and Levy, K.N. (2004). The influence of client variables on psychotherapy. In *Bergin & Garfield's handbook of psychotherapy and behavior change* (ed. M.J. Lambert), pp. 194–226. Wiley, New York.
93. Piper, W.E., Joyce, A.S., Azim, H.F.A., *et al.* (1994). Patient characteristics and success in day treatment. *Journal of Nervous and Mental Diseases*, **179**, 432–8.
94. Holmes, J. (2003). Borderline personality disorder and the search for meaning: an attachment perspective. *The Australian and New Zealand Journal of Psychiatry*, **37**(5), 524–31.
95. Fonagy, P., Leigh, T., Steele, M., *et al.* (1996). The relation of attachment status, psychiatric classification, and response to psychotherapy. *Journal of Consulting and Clinical Psychology*, **64**, 22–31.
96. Sandler, J., Dare, C. and Holder, A. (1992). *The patient and the analyst* (2nd edn). Karnac, London.
97. Limentani, A. (1972). The assessment of analysability: a major hazard in selection for psychoanalysis. *The International Journal of Psycho-analysis*, **53**, 351–61.
98. Kernberg, O., Clarkin, J.F., and Yeomans, F.E. (2002). *A primer of transference focused psychotherapy for the borderline patient*. Jason Aronson, New York.
99. Bateman, A.W. and Fonagy, P. (2004b). *Psychotherapy for borderline personality disorder: mentalization based treatment.* Oxford University Press, Oxford.
100. Winnicott, D.W. (1965). *The maturational process and the facilitating environment.* Hogarth Press, London.
101. Ryle, A. (2004). The contribution of cognitive analytic therapy to the treatment of borderline personality disorder. *Journal of Personality Disorders*, **18**(1), 3–35.
102. Gorman, H.E. (2002). Growing psychoanalysis. *Canadian Journal of Psychoanalysis*, **10**(1), 45–69.
103. Feldman, M. (1993). The dynamics of reassurance. *The International Journal of Psycho-analysis*, **74**, 275–85.
104. Sandler, J. and Sandler, A.M. (1994). Theoretical and technical comments on regression and anti-regression. *The International Journal of Psycho-analysis*, **75**, 431–9.
105. Inderbitzin, L.B. and Levy, S.T. (2000). Regression and psychoanalytic technique. *The Psychoanalytic Quarterly*, **69**(2), 195–223.
106. Smith, H.F. (1997). Resistance, enactment, and interpretation: a self-analytic study. *Psychoanalytic Inquiry*, **17**(1), 13–30.
107. Cairo-Chiarandini, I. (2001). To have and have not: clinical uses of envy. *Journal of the American Psychoanalytic Association*, **49**, 1391–404.
108. Smith, H.F. (2003a). Analysis of transference: a north American perspective. *The International Journal of Psycho-analysis*, **84**, 1017–41.
109. Kernberg, O.F. (1984). *Severe personality disorders: psychotherapeutic strategies*. Yale University Press, New Haven, CT.
110. Bowlby, J. (1988). *A secure base: clinical applications of attachment theory*. Routledge, London.
111. Diamond, D., Stovall-McClough, C., Clarkin, J.F., *et al.* (2003). Patient-therapist attachment in the treatment of borderline personality disorder. *Bulletin of the Menninger Clinic*, **67**(3), 227–59.
112. Farber, B.A., Lippert, R., and Nevas, D. (1995). The therapist as attachment figure. *Psychotherapy*, **32**, 204–12.
113. Mallincrodt, B., Porter, M., and Kivlighan, M. (2005). Client attachment to therapist, depth of in-session exploration and object relations in brief psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, **42**, 85–100.
114. Parish, M. and Eagle, M.N. (2003). Attachment to the therapist. *Psychoanalytic Psychology*, **20**(2), 271–86.
115. Bradley, R., Heim, A.K., and Westen, D. (2005). Transference patterns in the psychotherapy of personality disorders: empirical investigation. *The British Journal of Psychiatry*, **186**, 342–9.
116. Eames, V. and Roth, A. (2000). Patient attachment orientation and the early working alliance: a study of patient and therapist reports of alliance quality and ruptures. *Journal of Psycho-Therapy Research*, **10**, 421–34.
117. Waldinger, R.J., Seidman, E.L., Gerber, A.J., *et al.* (2003). Attachment and core relationship themes: wishes for autonomy and closeness in the narratives of securely and insecurely attached adults. *Psychotherapy Research*, **13**(1), 77–98.
118. Spence, D.P. (1982). *Narrative truth and historical truth. Meaning and interpretation in psychoanalysis*. Norton, New York/London.
119. Joseph, B. (1985). Transference: the total situation. *The International Journal of Psycho-analysis*, **66**, 447–54.
120. Cavell, M. (1994). *The psychoanalytic mind*. Harvard University Press, Cambridge, MA.

121. Hausner, R.S. (2000). The therapeutic and working alliances. *Journal of American Psychoanalytic Association*, **48**(1), 155–87.
122. Strachey, J. (1934). The nature of the therapeutic action of psychoanalysis. *The International Journal of Psycho-analysis*, **15**, 275–92.
123. Couch, A.S. (2002). Extra-transference interpretation. *Psychoanalytic Study of the Child*, **57**, 63–92.
124. Hoglend, P., Amlo, S., Marble, A., *et al.* (2006). Analysis of the patient-therapist relationship in dynamic psychotherapy: an experimental study of transference interpretations. *The American Journal of Psychiatry*, **163**(10), 1739–46.
125. Hoglend, P., Johansson, P., Marble, A., *et al.* (2007). Moderators of the effect of transference interpretation in brief dynamic psychotherapy. *Psychotherapy Research*, **17**(2), 162–74.
126. Bollas, C. (1994). Aspects of the erotic transference. *Psychoanalytic Inquiry*, **14**(4), 572–90.
127. Bowlby, J. (1977). The making and breaking of affectional bonds. II. Some principles of psychotherapy. *The British Journal of Psychiatry*, **130**, 421–31.
128. Fonagy, P., Gergely, G., Jurist, E., *et al.* (2002). *Affect regulation, mentalization and the development of the self*. The Other Press, New York.
129. Spillius, E.B. (1992). Clinical experiences of projective identification. In *Clinical lectures on Klein and Bion* (ed. R. Anderson), pp. 59–73. Routledge, London.
130. Freud, S. (1912b). Recommendations to physicians practising psychoanalysis. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 12 (ed. J. Strachey), pp. 109–120. Hogarth Press, London.
131. Heimann, P. (1950). On countertransference. *The International Journal of Psycho-analysis*, **31**, 81–4.
132. Renik, O. (1998). The analyst's subjectivity and the analyst's objectivity. *The International Journal of Psycho-analysis*, **79**, 487–97.
133. Mitchell, S.A. (1997). *Influence and autonomy in psychoanalysis*. Analytic Press, Hillsdale, N.J.
134. Racker, H. (1968). *Transference and countertransference*. Hogarth Press, London.
135. King, P. (1978). Affective response of the analyst to the patient's communications. *The International Journal of Psycho-analysis*, **59**, 329–34.
136. Sandler, J. (1993). Communication from patient to analyst: not everything is projective identification. *British Psycho-Analytical Society Bulletin*, **29**, 8–16.
137. Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, **8**(9), 396–403.
138. Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, **27**, 169–92.
139. Fonagy, P. and Target, M. (2007). The rooting of the mind in the body: new links between attachment theory and psychoanalytic thought. *Journal of the American Psychoanalytic Association*, **55**(2), 411–56.
140. Winnicott, D.W. (1949). Hate in the countertransference. *The International Journal of Psycho-analysis*, **30**, 69–75.
141. Betan, E., Heim, A.K., Zittel Conklin, C., *et al.* (2005). Countertransference phenomena and personality pathology in clinical practice: an empirical investigation. *The American Journal of Psychiatry*, **162**(5), 890–8.
142. Leuzinger-Bohleber, M., Stuhr, U., Ruger, B., *et al.* (2003a). How to study the 'quality of psychoanalytic treatments' and their long-term effects on patients' well-being: a representative, multi-perspective follow-up study. *The International Journal of Psycho-analysis*, **84**(Pt 2), 263–90.
143. Blum, H.P. (1989). The concept of termination and the evolution of psychoanalytic thought. Annual meeting of the American Psychoanalytic Association (1987, Montreal, Canada). *Journal of the American Psychoanalytic Association*, **37**, 275–95.
144. Bateman, A.W. and Fonagy, P. (2006). *Mentalization based treatment for borderline personality disorder: a practical guide*. Oxford University Press, Oxford.
145. Fonagy, P. (1999). Memory and therapeutic action (guest editorial). *The International Journal of Psycho-analysis*, **80**, 215–23.
146. Steiner, J. (1993). *Psychic retreats: pathological organisations in psychotic, neurotic and borderline patients*. Routledge, London.
147. Fonagy, P. and Bateman, A. (2006). Progress in the treatment of borderline personality disorder. *The British Journal of Psychiatry*, **188**, 1–3.
148. O'Shaughnessy, E. (1992). Enclaves and excursions. *The International Journal of Psycho-analysis*, **73**, 603–11.
149. Symington, N. (1983). The analyst's act of freedom as agent of therapeutic change. *International Review of Psycho-analysis*, **10**, 783–92.
150. Ehrenberg, D. (1993). *The intimate edge*. Norton, New York.
151. Safran, J.D. (2003). The relational turn, the therapeutic alliance, and psychotherapy research: strange bedfellows or postmodern marriage? *Contemporary Psychoanalysis*, **39**, 449–75.
152. Bateman, A.W. and Fonagy, P. (2004a). Mentalization-based treatment of BPD. *Journal of Personality Disorders*, **18**(1), 36–51.
153. Klein, M. (1950). On the criteria for the termination of a psychoanalysis. *The International Journal of Psycho-analysis*, **31**, 78–80.
154. Payne, S. (1950). Short communication on criteria for terminating analysis. *The International Journal of Psycho-analysis*, **31**, 205.
155. Pedder, J. (1988). Termination reconsidered. *The International Journal of Psycho-analysis*, **69**, 495–505.
156. Bachrach, H.M., Galatzer-Levy, R., Skolnikoff, A., *et al.* (1991). On the efficacy of psychoanalysis. *Journal of the American Psychoanalytic Association*, **39**, 871–916.
157. Fonagy, P., Kachele, H., Krause, R., *et al.* (2002). *An open door review of outcome studies in psychoanalysis* (2nd edn). International Psychoanalytical Association, London.
158. Lazar, S.G. (ed.) (1997). *Extended dynamic psychotherapy: making the case in an era of managed care*. Analytic Press, Hillsdale, NJ.
159. Leuzinger-Bohleber, M. and Target, M. (eds.) (2002). *The outcomes of psychoanalytic treatment*. Whurr, London.
160. Richardson, P., Kachele, H., and Renlund, C. (eds.) (2004). *Research on psychoanalytic psychotherapy with adults*. Karnac, London.
161. Stanton, A.H., Gunderson, J.G., Knapp, P.H., *et al.* (1984). Effects of psychotherapy in schizophrenia. I. Design and implementation of a controlled study. *Schizophrenia Bulletin*, **10**, 520–63.
162. Bateman, A.W. and Fonagy, P. (1999). The effectiveness of partial hospitalization in the treatment of borderline personality disorder-a randomised controlled trial. *The American Journal of Psychiatry*, **156**, 1563–9.
163. Bateman, A.W. and Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: an 18-month follow-up. *The American Journal of Psychiatry*, **158**(1), 36–42.
164. Blomberg, J., Lazar, A., and Sandell, R. (2001). Outcome of patients in long-term psychoanalytical treatments. First findings of the Stockholm outcome of psychotherapy and psychoanalysis (STOPP) study. *Psychotherapy Research*, **11**, 361–82.
165. Grant, J. and Sandell, R. (2004). Close family or mere neighbours? Some empirical data on the differences between psychoanalysis and psychotherapy. In *Research on psychoanalytic psychotherapy with adults* (eds. P. Richardson, H. Kächele, and C. Renlund), pp. 81–108. Karnac, London.
166. Sandell, R., Blomberg, J., Lazar, A., *et al.* (2000). Varieties of long-term outcome among patients in psychoanalysis and long-term psychotherapy: a review of findings in the Stockholm outcome of

psychoanalysis and psychotherapy project (STOPP). *The International Journal of Psycho-analysis*, **81**(5), 921–43.
167. Leuzinger-Bohleber, M., Stuhr, U., Ruger, B., *et al.* (2003b). How to study the quality of psychoanalytic treatments and their long-term effects on patients' well-being: a representative, multi-perspective follow-up study. *The International Journal of Psycho-analysis*, **84**, 263–90.
168. Gabbard, G.O., Gunderson, J.G., and Fonagy, P. (2002). The place of psychoanalytic treatments within psychiatry. *Archives of General Psychiatry*, **59**(6), 505–10.
169. Fonagy, P. (2006). Evidence-based psychodynamic psychotherapies. In *Psychodynamic diagnostic manual* (ed. PDM Task Force). Alliance of Psychoanalytic Organizations, Silver Spring, MD.
170. Bachrach, H.M., Weber, J.J., and Murray, S. (1985). Factors associated with the outcome of psychoanalysis. Report of the Columbia psychoanalytic research center (IV). *International Review of Psychoanalysis*, **12**, 379–89.
171. Erle, J. and Goldberg, D. (1984). Observations on assessment of analyzability by experienced analysts. *Journal of the American Psychoanalytical Association*, **32**, 715–37.
172. Sandell, R., Blomberg, J., Lazar, A., *et al.* (1997). *Findings of the Stockholm outcome of psychotherapy and psychoanalysis project* (*STOPPP*). Paper presented at the Annual Meeting of the Society for Psychotherapy Research, Geilo, Norway.
173. Dührssen, A. (1962). Katamnestische Ergebnisse bei 1004 Patienten nach analytischer Psychotherapie. *Psychosomatic Medicine*, **8**, 94–113.
174. Breyer, F., Heinzel, R., and Klein, T. (1997). Kosten und Nutzen ambulanter Psychoanalyse in Deutschland (Cost and benefits of outpatient psychoanalytic therapy in Germany). *Gesundheitsökonomie und Qualitätsmanagement*, **2**, 59–73.
175. Bateman, A.W. and Fonagy, P. (2003). Health service utilization costs for borderline personality disorder patients treated with psychoanalytically oriented partial hospitalization versus general psychiatric care. *The American Journal of Psychiatry*, **160**(1), 169–71.
176. Clarkin, J., Levy, K.N., Lenzenweger, M.F., *et al.* (2007). Evaluating three treatments for borderline personality disorder: a multiwave study. *The American Journal of Psychiatry*, **164**, 922–8.
177. Giesen-Bloo, J., van Dyck, R., Spinhoven, P., *et al.* (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, **63**(6), 649–58.

6.3.6 Group methods in adult psychiatry

John Schlapobersky and Malcolm Pines

Introduction

After a century of development, group therapy is today one of the most widely practised treatment methods in psychiatry with an extensive literature. There are three principles common to its wide range of applications. First, the therapist calls the 'community' into the consulting room where, together with the therapist, it becomes the therapeutic agent. Second, the therapist assembles a group of people who can contribute to a commonly held resource from which its members can each derive benefits. And third, the therapist does nothing for them in the context of the group, that they can do for themselves, and one another.

This chapter starts by providing a conceptual framework that differentiates methods, models, and applications for the practice of group therapy in adult psychiatry. After classification of the different methods and applications we discuss the main theoretical models; explore the dynamic life of therapy groups; consider some of the key clinical issues facing practitioners; their applications to a range of patient populations and settings; their evaluation and justification and their historical evolution this century. In the conclusion we consider the planning of group services and the training of their practitioners. This revision of the chapter has brought it up-to-date with the contemporary literature in a field that has seen a great deal of innovation since the original 2000 edition.

The developing evidence base for group psychotherapy is 'Guardedly optimistic. The literature has become stronger and deeper and is capable of supporting evidence-based treatment recommendations for some patient populations.' The evidence base for the effectiveness of group psychotherapy has been growing with the field. Some 700 studies, spanning the past three decades, have shown that the group format consistently produced positive effects with diverse disorders and treatment models.[1] These show that both individual and group psychotherapy will effect much the same set of results. For group therapy to be effective it has to utilize those therapeutic factors originally laid out by Foulkes[2] and later by Yalom[3]—the group has to be the primary focus of therapy; patients need to be well selected; and therapists need to be adequately trained. The chapter will address these questions of focus, selection, and training.

Although the two authors of this chapter are both group analysts, we have set out to provide a full account of the wide range of group work practice. The United Kingdom is our own working location which lends emphasis to the chapter but it is compiled with sources and references that address the international field and it gives attention to current literature in many countries including North and South America and Continental Europe.

Basic methods

In Fig. 6.3.6.1 we have used two simple factors—therapeutic goals and group leadership—to provide a simple classification of the many different methods.

Therapeutic goals

Groups will be more or less specific in their therapeutic goals. For example, those catering for a homogeneous population with a commonly defined problem whose solution provides the basis for entry to the group—such as overcoming drink or drug dependence—are classified here with specific goals. Groups that provide psychoanalytic psychotherapy, whether run according to Interpersonal, Tavistock, or Group-Analytic models, are classified here with non-specific goals. There is a wide range of variation between these extremes and within each of these main psychodynamic models.

Leadership

The more the leader directs the group, the more prominent he becomes as the group's 'model object'. The less the leader directs the group, the more scope there is for the emergence of unconscious dynamics and for attention to transference and counter-transference. In this case, therapy progresses through the development of relationships. The greater the leadership activity, the more likely it

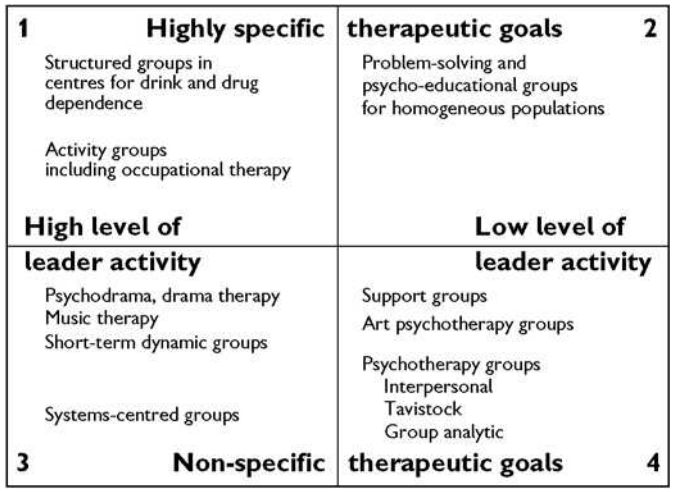


Fig. 6.3.6.1 A simple classification of group methods.

is that group members are being offered a technique or skill in the setting of a group. The lower this activity, the stronger will be the relational content of the therapy and the therapist's skills of fostering relationships will equip the group to work developmentally and in depth on both the obvious and hidden issues that its members bring. The three principal psychodynamic methods discussed in The principal model of psychodynamic group therapy below share a non-directive philosophy with subtle but significant differences between their models of practice.

Using these two basic indicators—specificity of goals and levels of leadership activity—the four quadrants in the diagram provide us with a simple way of 'placing' the different group therapies.

Goal-specific therapy with high level of leader activity: quadrant 1

In many drug and alcohol dependency regimes, participants are required to fulfil obligations tied to each stage of a structured programme. They move forward when stage-specific obligations are fulfilled. As the novice moves up, he/she becomes a trainer to the newcomers with the therapist(s) directing the process in active terms. Cognitive therapy given in a group setting uses the group as an assembly who learn from and discuss with the expert. Dependency on a shared and valued leader, and attention to group dynamics amplify the learning and some group cohesion develops, but this is not the primary focus of the therapy.

Goal-directed with lower level of leader activity: quadrant 2

Problem-solving or psycho-educational groups for homogeneous populations, such as those set up for eating disorders or offenders, which are run along analytic lines, can be placed in this category. Although there are clear and directed goals, the leader's level of activity is confined to a facilitating, linking, or enabling one, followed by analysis and interpretation. Group discussion and cohesion amplify the affective experience and enhance the learning.

Non-specific goals with high level of leader activity: quadrant 3

In **psychodrama groups**, leadership is explicitly vested in the psychodrama director. The needs and goals of group members are diffuse and often diverse and, in psychiatric practice, will have to do with relief from mental suffering. The psychodrama director can draw on many techniques. Affective arousal can be high so the power of sharing through discussion and the sympathy; and empathy of group members towards one another become powerful therapeutic tools. Strong conflict arousal and its subsequent resolution is similarly therapeutic.

Systems-centred therapy (as developed by Agazarian)[4] similarly provides a high level of leadership activity for groups that have non-specific goals. Short-term dynamic groups are frequently constituted with non-specific goals but are run over 10 or 20 sessions by leaders who maintain a high level of involvement and direction, often demarcated according to the different stages of the group's progress.

Non-specific goals with low level of leader activity: quadrant 4

The goals of group-analytic or psychoanalytic group therapy are most frequently diffuse and non-specific involving relief from symptoms and other forms of suffering; personal growth; and psychological change. There are three main schools considered in this chapter—the Interpersonal (Yalom), Tavistock (Bion or group-as-a-whole), and the Group-Analytic. They share non-specific goals and have low levels of leader activity but differ from one another in how the leadership role and function is understood and discharged. They share assumptions about the importance of unconscious individual and group dynamics and look to the group for its transformational potential. Their differences affect the way in which transference and counter-transference is understood and worked with. There is a comparative appraisal of these models below.

The application of basic methods

We now provide a more detailed overview of the current field and offer a brief set of training requirements for practitioners in each of the methods discussed.

The field can be divided into **five basic methods**—activity, supportive, problem-solving, psycho-educational, and psychodynamic. The first three methods are goal-specific as indicated by their descriptions, the fourth is less specific and the fifth is a non-directive analytic psychotherapy. In supportive and problem-solving groups, therapeutic leadership can be highly directed or not, depending on the approach. Activity and psycho-educational groups will inevitably have a high level of directed group leadership, whilst psychodynamic groups have a much lower level of directed group leadership. All five methods rely on the same basic procedures—the selection and grouping of a number of people seeking help who have regular meetings together with one or more well-trained therapist(s).

Activity groups

The most vulnerable and disturbed patients can be placed in therapy groups defined by an activity that provides a convening function such as exercise or cooking. They can then be used to create conditions for a wide range of secondary functions that foster affiliations, develop social skills, address unspoken anxieties, and express troubling emotions. Occupational therapists and nurses using art media or other socially syntonic activity like gardening or

hair-dressing have been developing a wide range of group services in both acute and rehabilitation psychiatry for many years.[5] The approach has been used in a wide range of other settings including medical rehabilitation, rehabilitation with refugees, social work and fostering, and adoption programmes. Groups that keep the original activity as their primary focus, working with art media for example, need to be differentiated from those which use such media to develop an analytic focus on psychological work. The arts psychotherapies belong to this latter group. They have non-specific therapeutic goals and might, as in the case of music therapy, have a high level of leadership activity or, in the case of art therapy, have a low level of leadership activity.

Whilst therapists do not engage in the uncovering and exploration of unconscious dynamics, they will need leadership abilities, capabilities in organizing group activities, and should have a basic understanding of psychopathology and group dynamics.

Supportive groups

These groups function as a form of social support providing containment, the improvement of social skills, and the enhancement of participants' capacities for social adaptation. They aim to reduce the deleterious effects of social isolation, bring people out of withdrawal into a social context, and provide opportunities for problem sharing. They cater to patient populations with long-standing personality disorders not open to uncovering exploration; those with chronic mental and physical illness,[6] physical handicap, mental retardation, and carers for those with any of these problems. They will often allow a certain amount of psycho-education with the group leader influencing members' attitudes as in the case, for example, of a group for young sexually active adults with learning disability who might receive guidance on contraception.

Whilst therapists do not engage in the uncovering and exploration of unconscious dynamics, they will need leadership abilities, capabilities in organizing group activities, and should have a basic understanding of psychopathology and group dynamics.

Problem-solving groups

Group therapy is provided for a set of referral criteria to resolve a defined and sometimes circumscribed problem. Alcoholics Anonymous, Alanon, Gamblers Anonymous, and groups for people with poor impulse control, eating disorders, or other habitual problems such as smoking, are a few of the examples. These groups can take on many of the features of long-term support groups, in that they offer ego-supportive and adaptive resources, providing an extended service for monitoring by the patient or by professionals, without necessarily committing members to the deeper and more radical analytic work entailed by a psychodynamic group. In many cases, the problem-solving focus provides a convening frame by which to engage a population who are soon drawn into psychodynamic work that sees them through profound changes. Many of the groups run by clinicians in primary hospital care—occupational therapists, nurses, doctors, and psychologists—take this form. The Group Work Programme at the Medical Foundation For Victims Of Torture in London is another example (see also section on trauma in specical population, below).

Where they cater for the more severely disturbed, staff will need to be well-trained in one of the core professions. They will need leadership abilities, capabilities in organizing group activities, should have a basic understanding of psychopathology, and be sufficiently well-trained to explore the dynamic group issues that lead from the problem back to the personality structure of their membership. If therapeutic goals involve major changes in personality and social functioning, this will involve the uncovering and exploration of unconscious dynamics.

Psycho-educational groups

The original groups for servicemen with war neurosis at Northfield took this form in which people were given the role of students of their disorders rather than 'sick' people. Patients become open to new information and are better able to unlearn maladaptive attitudes about the nature of their disorder. This more cognitive approach can be applied in homogeneous problem-solving groups. Information can be provided through lectures, discussions, and suitable reading material. There are different ways of lowering anxiety and uncovering maladaptive and inappropriate attitudes towards such problems as anxiety states, phobias and obsessions, and psychosomatic disorders. Many of the groups run for those with serious physical illness (see also section on the medically ill in special populations, below) take this form. And there is often a major psycho-educational component in support groups—for example, those with chronic mental illness who can be helped to understand and cope with delusions, hallucinations, and the stigma of illness.[7] (see also section on the mentally ill in special populations, below)

Staff need leadership abilities; capabilities in organizing group activities, a basic understanding of psychopathology, and need to be sufficiently well-trained in their chosen problem area to be able to relate its educational focus to thematic group issues. If therapeutic goals involve major changes in personality and social functioning, this will involve the uncovering and exploration of unconscious dynamics.

Psychodynamic groups

There are supportive, problem-solving, and psycho-educational components in all psychodynamic groups, but the description 'psychodynamic' is reserved for those in which the declared goal is lasting personal change through a non-directive, free-associative therapy. The range of different group contexts is so varied that—at first sight—they might appear to have little in common. But there will be common principles offering therapy to a group of people on an in-patient unit recovering from psychosis and meeting thrice weekly; those in a secure unit for violent offenders meeting once weekly; and those—including mental health trainees—attending a group in private practice once or twice weekly. These principles can be summarized in the table, Table 6.3.6.1 below.

Within these parameters therapy is part of the cultural domain of all shared, conversational experience in which people struggle with meaning—in congregational life, in the confessional, in theatre, narrative or poetry.[8,9]

Staff need to be trained to the level already described. They need good leadership ability, capabilities in organizing group activities, and a good understanding of psychopathology. Beyond these requirements, therapeutic goals will involve major changes in personality and social functioning involving the uncovering and exploration of unconscious dynamics. So staff will need access to a

Table 6.3.6.1 Organizing principles for group therapy

1	Members will have been chosen by the therapist.
2	They will have chosen to join and participate.
3	They will be expected to justify their place by reliable attendance and participation.
4	The work will be governed by a psychotherapeutic contract.
5	The contract will include definitions of confidentiality and other boundaries.
6	A therapeutic alliance with individual members will be established either prior to their joining the group or during the early stages of their attendance in the group.
7	Agreed parameters will include the duration and time-boundaries of the group as well as its membership and composition.
8	Groups can be homogeneous or heterogeneous.
9	Groups may have a fixed time limit or continue on a slow–open basis for many years.
10	Groups may have a stable and fixed, or a rotating membership with empty places taken by new members.

range of specialized training opportunities which should provide psychodynamic theory, clinical supervision and, ideally, some opportunity for the practitioners' own personal development.

The contemporary field and its history

The paradigm shift that led to a vision of the group as a whole began over a period of time and in a number of locations. Trigant Burrow coined the term group analysis in the USA in the 1920s.[10,11] Further development in the years after World War II period enabled workers to recognize the dynamics of groups and institutions and led to group and family therapy, milieu therapy, and therapeutic community concept and practice.

Second World War

In Britain and the US, during the Second World War, an appreciation of group psychology led to a wide range of innovations, the most important of which included:

- The use of group methods for selection and allocation of work responsibilities
- Studies of group morale
- The integration of psychiatric knowledge to the management of large groups through the role of the command psychiatrist
- The treatment of acute and prolonged battle stress and the rehabilitation of returned prisoners of war.

Clinicians used the opportunities created within army psychiatry to apply methods developed in the pre-war years.[12] Brigadier J.R. Rees, Director of the Tavistock Clinic in the 1930s, largely created this opportunity in the British Army. His Tavistock colleagues formed an 'invisible college' and were responsible for signal achievements in the advances in selection and treatment.[13]

Foremost amongst these were Northfield, the military hospital near Birmingham where S.H. Foulkes was a senior medical officer.[14] A refugee from Nazi Germany, Foulkes brought with him from the Frankfurt Institute the revised understanding of Freudian theory that was also to prove influential in the US. In the New School for Social Research, New York, and in the work of Neo-Freudians like Erich Fromm, Frieda Fromm-Reichman, and social theorists like Adorno, Marcuse, and Norbert Elias, psychoanalysis and Marxist theory were brought into a new, creative relationship.[15] In the United Kingdom Foulkes first developed his approach to group therapy in Exeter before the war and was able to apply it successfully on a large scale to the treatment of war neuroses at Northfield.

Post-war period

Group psychotherapy moved from inspirational and didactic models to psychodynamic and analytic ones in the post-war period.

(a) United Kingdom

The Group-Analytic Tradition: Foulkes gathered around him a small group of clinicians and others who developed his ideas and practises. Drawing on the ideas of Trigant Burrow, they called it group analysis and later established the Group-Analytic Society, and trained generations of clinicians. His first book written in the heat of the Northfield experience outlined the basics of his approach.[16] Other publications followed and, with Malcolm Pines, training courses were established which lead to the founding of the Institutes of Group Analysis and Family Therapy, the Association of Family Therapists, and Association of Therapeutic Communities. There are now training courses in group analysis in many centres in the United Kingdom and continental Europe. The Journal, *Group Analysis*, established by Foulkes, continues to be the major publication in European group psychotherapy. Group-analytic psychotherapy has undergone clinical evaluation by a number of clinicians.[17–19]

The 'Tavistock' Approach: The approach originates in the work of Bion, Ezriel, Sutherland, and their colleagues. It shares with group analysis an interest in the underlying pattern of object relations in groups but, under the influence of Bion—its major exponent—his 'basic assumption theory' is applied to the exclusion of almost everything else. (Bion's monograph was his only publication on groups and marked the end of his interest in the subject.[20]) The approach has been especially influential in staff training and consultancy which, given the slender theoretical foundations on which it rests, suggests a wide responsiveness in the field of basic assumption theory. The approach has undergone further development in the United States where it is often referred to as group-as-a-whole. When employed as a therapy, it can overlook the individuality of a group's members, disturbing some patients whose experience of the group situation can repeat early developmental traumas of neglect and misunderstanding by caretakers. Malan's study of effectiveness[21] raised serious questions about the model's efficacy in its clinical applications but its training applications continue to influence the field.

(b) United States

(i) Early pioneers

Jacob Moreno was the innovator of group psychodrama, a pioneer form of group psychotherapy.[22] He also introduced sociometry, a scientific method for the study of group affiliations and conflicts, widely accepted and used by social psychologists. **Slavson** was an educationalist of psychoanalytic persuasion who became the central figure in the early development of group psychotherapy. His clinical influence, particularly with groups for the parents of children in difficulty, and his focus on the dynamics of projection in groups has been of lasting importance.[23] His organizational efforts lead to the formation of the American Group Psychotherapy

Association. **Emanuel Schwartz** began to apply psychoanalytic ideas to group psychotherapy in the late 1930s and was later joined by **Alexander Wolf**.[24, 25] In their approach, people underwent an individual psychotherapy in the setting of a group, a kind of parallel process alongside their fellow patients, with attention focused on the transferential relationship between each individual and their therapist. The approach has been of lasting importance in creating a clinical framework for combining individual and group therapy. Foulkes' criticism at the time was that the approach overlooked any systematic use of group-specific process. In contrast to their 'psychotherapy in the group' he offered group analysis as a clinical alternative, describing it as 'psychotherapy by the group'.

(ii) Irving Yalom

Yalom's interpersonal approach is influenced by the interpersonal psychotherapy of Sullivan and Frank. His *Theory And Practice of Group Psychotherapy,* now in its fifth edition and written jointly with Leszcz, is the first systematic account of groups informed by research and remains one of the most influential books in the field.[3] Yalom's later text on inpatient group psychotherapy systematized group work in that setting.[26]

(iii) The contemporary field

There are many centres of excellence, a wide range of methods and models and an empirical base grounded in research. The most useful single text is by Rutan and Stone, now in its fourth edition.[27] Collections by Kaplan and Sadock[28] and by Alonso and Swiller[29] cover the field. Psychoanalytic models have a rich diversity of theory with contributions from object relations, self psychology, and social systems theory.[30] The Modern Group movement is amongst the most innovative, beginning with a classic text by Spotnitz[31] and developing through an active training programme, a journal, *The Modern Group,* and publications by Ormont[32] and others.

(iv) South America

There is a vigorous field of development throughout South America that draws on both the Tavistock and Group Analytic traditions but is informed by independent sources based largely on the work of Pichon-Riviere. Tubert-Oklander and Hernandez de Tubert have introduced this approach, referred to as Operative Groups, to the English-speaking world.[33]

(v) Continental Europe

Group methods have played an active part in the reconstruction of mental health services throughout Europe in the post-war period. Distinctive approaches are emerging. Those in Germany include psychosomatic practice[34] and the Gottingen model.[35] The journal of the Heidelberg Institute of Group Analysis gives access to a vigorous field. A major research study is in process, based in Germany, in which therapists throughout Europe are taking part and which aims to provide a detailed evaluation of group therapy, its patients, and its therapists.[36] Other distinctive developments include those in Italy[43] and original training models, for example, those in Greece[37] and Norway.[38]

Principal models of psychodynamic group therapy

The therapist

The therapist is responsible to the group— and to the institution in which it is set—for achieving and maintaining professional competence and should have a level of training appropriate to the task. A formal qualification in psychotherapy is the ideal training. This will have included theory, personal therapy for the therapist, and clinical supervision. Mental health professionals from all disciplines make an active contribution to a rich and diverse service with the training requirements of theory and supervision arranged at their workplace. The opportunity to run a group is provided in most psychiatric and psychology training programmes. Many centres and training institutions offer training in group methods and several are wholly committed to the training of group therapists who have their own professional associations in the UK and internationally. Private once- or twice-weekly analytic groups are now regarded by many mental health professionals as the therapy of choice for their personal development. Other requirements for a good therapist are listed in table 6.3.6.2.

(a) Making a beginning

The establishment of a group begins as a management task in the definition of its goals, recruitment of its members, protection of its setting, venue and timetable, and in the maintenance of its ongoing life. It evolves as a therapeutic task in which the therapist is responsible for maintaining a therapeutic attitude to the individual members and to the group as a whole. Powerful affects and attitudes will be directed towards her which she will monitor and transform into verbal and non-verbal therapeutic responses.

The therapeutic rationale will allow the therapist to be discriminating and consistent about interventions of various kinds during the life of the group and what follows below provides an orientation—based on the dynamic elements of structure, process and content—to the three main models used in the UK.

Structure, process, and content: the dynamic elements of a group

Regardless of the therapist's method, people usually start in groups with a form of serial monologue. Out of this arise the capacities to talk and listen that are often undeveloped or even non-existent at the outset of therapy but which are its core constituents. From talking and listening comes self-disclosure and out of this social exchange identification emerges, which in due course leads to dialogue and differentiation. So the conductor must give a place to monologue whilst, at the same time, cultivating dialogue—the exchange between members or sub-groups—and, ultimately,

Table 6.3.6.2 Requirements for therapeutic competence

Requirements for therapeutic competence include:
1 The ability to follow complex interactions and processes.
2 The ability to discriminate between appropriate activity and a containing form of silence.
3 A reflective attitude and the capacity to consider and reflect upon the processes concerning both the individual members and the group as a whole.
4 An eye for both the visible and invisible group and a curiosity about the unconscious or otherwise hidden aspects of a group's life. This will require the therapist's access to their own internal process and a capacity to make use of it.
5 A therapeutic rationale for action related to the group tasks and leadership requirements including the psychopathology of the individuals, their psychodynamics, and group dynamics.

promoting discourse, defined here as the free interaction of participants in the flexible and complex exchange that distinguishes the communication of a group.

Structure describes the more enduring aspects of any group's makeup, the 'architecture' of its interpersonal relations conceptualized first in terms of the setting and its boundaries and then conceptualized in the bond between each individual, the therapist(s), and the group as a whole. Process describes the fluid and dynamic fluctuations of emotion and experience, the business of relating and communicating, the changes of association and inter-member responses. The content of a group's exchange is in its visible and audible events, in the narrative line and dramatic content of peoples' encounters, the topics raised, their thematic development, and the extent to which they are explored or avoided.

As Fig. 6.3.6.2 illustrates, each of these three dynamic elements has a determining influence on each of the others. For example, a group in which there was a problem caused by the institution's failure to honour its commitment to reliable space for regular meetings, would have a serious structural problem. Intrusion, relocation, or a conflict over space might then emerge in the content of the members' associations as they talked about shared past experience. The therapist would need to decide whether to direct the process towards the connection between past and present anxieties, or reassurance that the therapist would—from now on—be able to protect their space.

(a) Overview of the Interpersonal, Tavistock, and Group-Analytic Models

In the Interpersonal school, intra-group interactions, including those between its members and the leader, are taken in their totality, but differentiating the leader as a different 'sort' of person from the others. In the Tavistock model, a two-body psychology is used to analyse the interchange between the leader and the group taken as a whole. The therapist's principal role is in the analysis and interpretation of defences against primitive anxieties (or basic assumptions). The Group-Analytic Model calls on elements of both foregoing models. Like the Tavistock model it considers the leader as structurally different to other group members but like the Interpersonal model, it encourages the leader to work in the group with individuals. A three-body psychology is used to understand the role of the leader who is referred to here as the conductor. Therapy proceeds through the dynamic interaction between each individual, the conductor, and the group as a whole.

(b) The model of interpersonal group therapy

The focus is on interpersonal learning as a primary mechanism of change. The group provides the antidote to maladaptive interpersonal beliefs and behaviours through feedback from others and encouragement to experiment with healthier behaviours first within the group and then outside. The joint examination of intra-group transference reactions allows members to replace processes that have a historical origin in the 'there and then', the dynamic past, with those more appropriate to the 'here and now', the dynamic present. The approach emphasizes the educational opportunities of working in the 'here and now' of the group. The therapist takes the responsibility for leading the group towards awareness of these interpersonal dynamics and their expressions. There is also greater therapist transparency than in other psychodynamic approaches with the therapist modelling desired behaviours, sharing the reactions to events in the group directly, and being open to feedback from other group members.

As the diagram indicates, interpersonal dynamics are kept at the forefront of members' attention by the therapist. This sets a pattern in which the content of members' discussions and the process of their interactions gives the group its agenda. The interpersonal approach places the therapist amongst other members of the group without giving him a distinctive structural identity and omits any formal demarcation for the boundaries of the group as a whole.

The model provided the early descriptive research on the phases of small groups, on the basis of which Yalom tabulated the **curative factors** in a group's life (see Table 6.3.6.3).

This construction has been very influential. Yalom's use of the term 'curative' poses many problems for those clinicians who see the goals of therapy involving personal growth and change. He did much to address this difficulty by singling out the last of his 'curative' factors—existential issues—for special treatment in a subsequent text.[49]

(c) The Tavistock model

Bion's ideas have an explanatory power and simplicity of application that continues to prove illuminating.[40] In a group at any

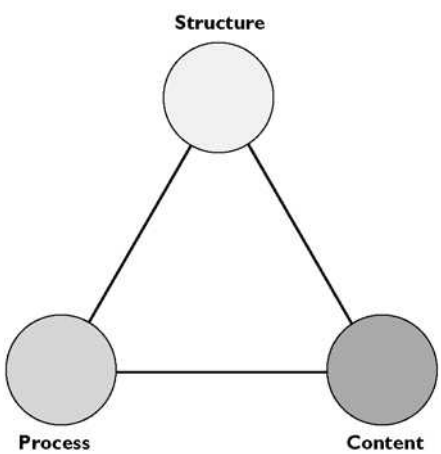


Fig. 6.3.6.2 The dynamic elements of a group.

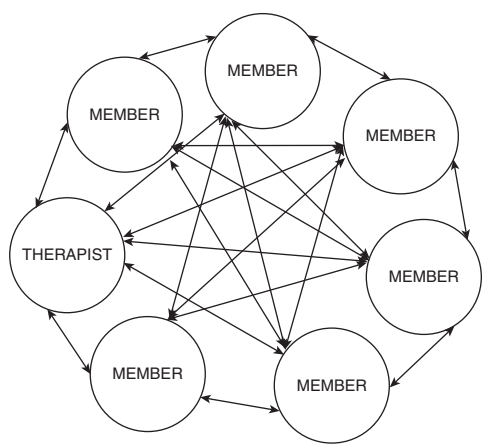


Fig. 6.3.6.3 The elements of an interpersonal group.

Table 6.3.6.3 Yalom's curative factors

1	Instillation of hope
2	Universality
3	Imparting information
4	Altruism
5	Corrective recapitulation of primary family group
6	Development of socializing techniques
7	Imitative behaviour
8	Interpersonal learning
9	Group cohesiveness
10	Catharsis
11	Existential factors

point in time, its culture and climate are governed by primitive, unconscious anxieties that impede its capacities for rational work in which the person or representation of the leader plays a crucial part. The anxieties, organized into one of three categories, referred to as **basic assumptions**, are dependency, fight or flight, and pairing. They affect the group as a whole in which only one basic assumption is believed to be operative at any point in time. Bion saw basic assumptions as interfering with the 'work group', the more rational, higher-level functioning of the group and its members. The therapist's key task lies in understanding and interpreting the operative basic assumption to the whole group. The meaning of individuals' experience is subsumed by this understanding of the whole. This therapist-centred approach sees transference only as directed towards the therapist who represents authority. In **dependency**, the group tries to elicit protection through passive or dependent behaviour. In **fight/flight** they will attack the therapist or some other issue; or retreat and withdraw. And in **pairing** they may create a group illusion that some magical form of rescue may arise from the dilemmas of group life through charged partnerships. Hopper has introduced a fourth basic assumption that he calls **massification/aggregation** in which the defensive structures of groups or societies in crisis is thought to entail either a rigid fusion of identities excluding individuality, or extensive withdrawal preventing mutuality.

The two-body psychology used here enforces a series of clinical constraints that reduce the complexity of group interaction to a bi-personal exchange between therapist and group taken as a whole. As Fig. 6.3.6.4 illustrates, intra-group dynamics are considered only in their entirety for what they reveal about the unconscious state of the group as a whole, and for what they indicate about the nature of the group's relationship with the therapist. Figure 6.3.6.4 illustrates how the therapist stands outside the group in a stance that is not only neutral and dispassionate but also opaque and withholding of self.

Ezriel, basing his work on Bion, developed his theory of **common group tension**.[41] He believed that the group would be caught up at any given time in a commonly shared conflict centred on the unconscious fear of catastrophe, what he called **the dreaded state**. People would avoid a state in the group—say one in which they talked about sad feelings—because of the unconscious fear that talking about sadness would lead to a dreaded state, in this case a depressive collapse. A group would be driven into unconscious, defensive organization—what he called **the required state**—to keep sadness at bay. For example, an extended period of manic humour, the required state, would help prevent **the avoided state**, sadness, and this would in turn protect against the dreaded state. Interpretations would allow members to become increasingly aware of the underlying catastrophic fears and reduce their need for defensive organization.

Horwitz calls Ezriel's approach 'deductive', in that it relates individual's contributions only to the common group tension. He realized that this deductive approach was clinically unproductive and developed, what he called an 'inductive' method which is group-centred.[42] Interventions are first addressed to individual members in the group. Only after working with patients individually does the therapist introduce a common theme that binds them together. Thus the therapist, as in the Group-Analytic model, works on a figure/ground basis in which individual contributions are valued and explored in their own right before they are contextualized in the life of the group as a whole.

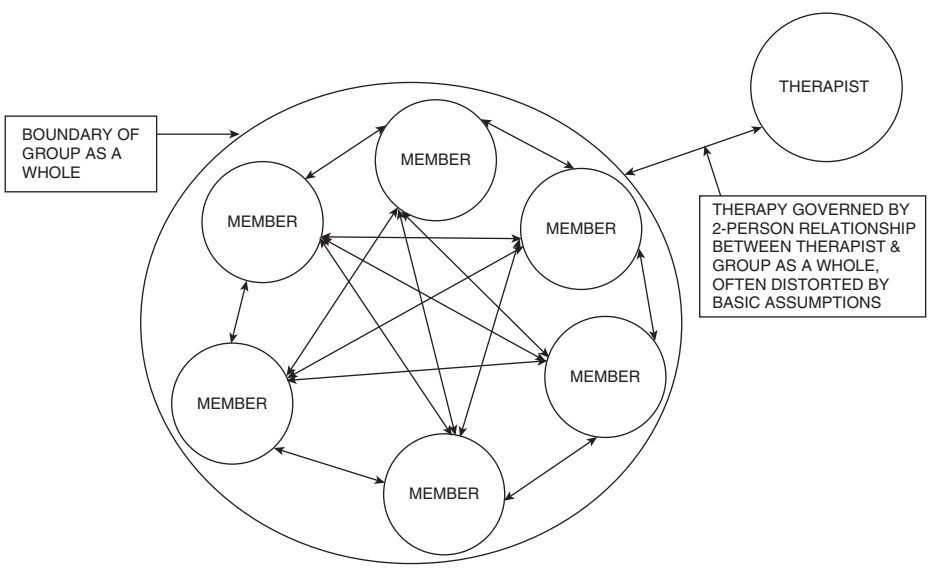


Fig. 6.3.6.4 The elements of a Tavistock group.

Another approach, **focal conflict theory**, as developed by Whitaker and Lieberman[43] is similar to Ezriel's in focussing the therapist on a conflict that becomes his point of emphasis, but it answers Horwitz's criticisms. On their account, underlying **disturbing motives** in group behaviour are acted against by unconscious, **restrictive solutions**. On their account the therapist—by focusing on such key conflicts—helps give members access to the unconscious anxieties and once these have been relieved they can construct more **enabling solutions** to the shared dilemmas of the group. The idea of focal conflicts, conceived of in these broader terms, has become an integral part of the Group-Analytic model.

(d) The Group-Analytic model

This approach integrates important aspects of the two preceding models but introduces a number of new elements. As Fig. 6.3.6.5 suggests, the therapist is encouraged to address the individual as well as the whole group and considers the more conscious and individual dynamics as well as the unconscious and potentially destructive whole-group dynamics. The approach is guided by an integrated set of concepts relating structure, process, and content to one another in which the group conductor works both as therapist and as group member to foster and cultivate the ordinary language of shared conversational experience. He will at times take up the position of the group's manager, and at other times he will speak personally as one of its members. Groups may begin with a relatively high level of leadership activity, referred to as **dynamic administration**, which is flexibly reduced with a decrescendo of responsibility as the group becomes the therapist and the leadership function is devolved upon its membership who becomes active co-therapists in each other's treatment[44]. Figure 6.3.6.6 indicates how, in this approach, at one key moment in the group its theme can focus on structural dynamics that link one member to both the therapist and to the group as a whole. The web of interconnecting dynamics between any one member and all the others, is summarized by 'the group as a whole', and this is called on to represent the inter-connecting latticework of relationships that includes all the members and the therapist.

(c) The Matrix

As Fig. 6.3.6.6 indicates, the conductor is inside and a part of the group, the structural elements of which provide a way of understanding the crucial links between each member, the conductor and the group as a whole. The triangle by which the group's psychological objects are linked to each other illustrates one of the 6 corresponding patterns of connection. When replicated for each of the 6 members, the diagram will produce a **matrix** of relational patterns, a complex relational field that will undergo change in terms of alliances, sub-groups, and polarizations. This concept of the matrix is crucial in group analytic theory. It allows us to accept that all events in a group will become part of an unconscious network that is intrapsychic, interpersonal, and transpersonal. The developing matrix creates the capacity to receive, contain, and eventually transform individuals' contributions, fostering integration at the individual level as it does so in the group as a whole.

Free-floating discussion: Free floating discussion is the group-analytic equivalent of free-association. The term originates in Foulkes' own writing and describes a set of key clinical concepts in therapeutic practice that distinguish the group-analytic approach. The language of the group is discussed in the dynamic life of groups, below.

Group-specific process: These processes, also mapped out originally by Foulkes, have been studied further by Pines[45] and by Agazarian and Peters.[46] The key concept of **resonance** describes the unconscious communication of emotion. The group provides its members with a wide field of meaning which is explored as they **mirror** one another's experience, find their emotions **amplified** by association with one another, and find **condensed**, sometimes highly aroused, cathartic experience, moments charged with significance.

Content analysis: Foulkes described four levels at which the content of the group's discussion can be analysed in the search for meaning.[47] In the Tavistock model this is the therapist's exclusive task, whereas here interpretation is only one amongst a number of others. The therapist's overall stance is to foster communication and educate the group's members about the dynamic links between the group's structure, the content of the discussion, and the form

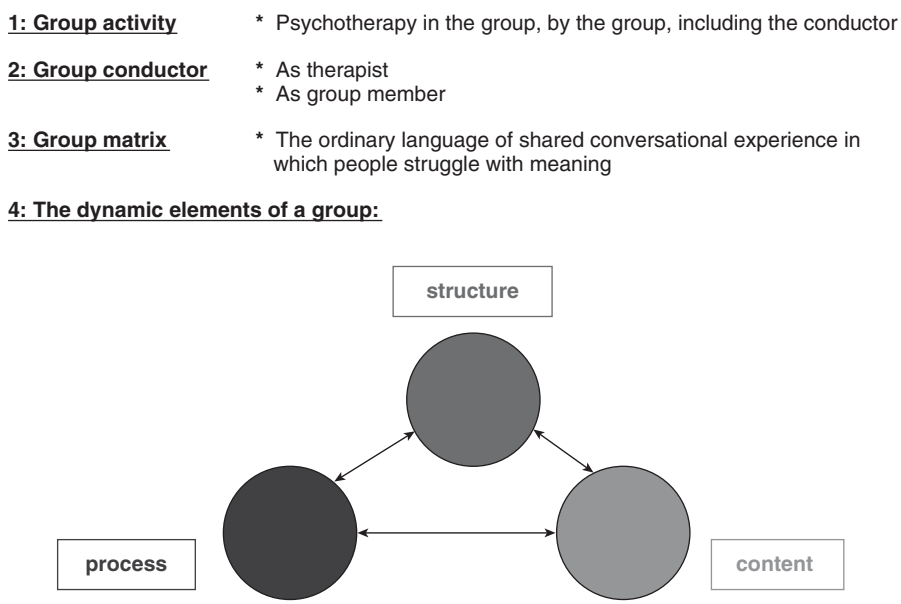


Fig. 6.3.6.5 Group Analytic psychotherapy.

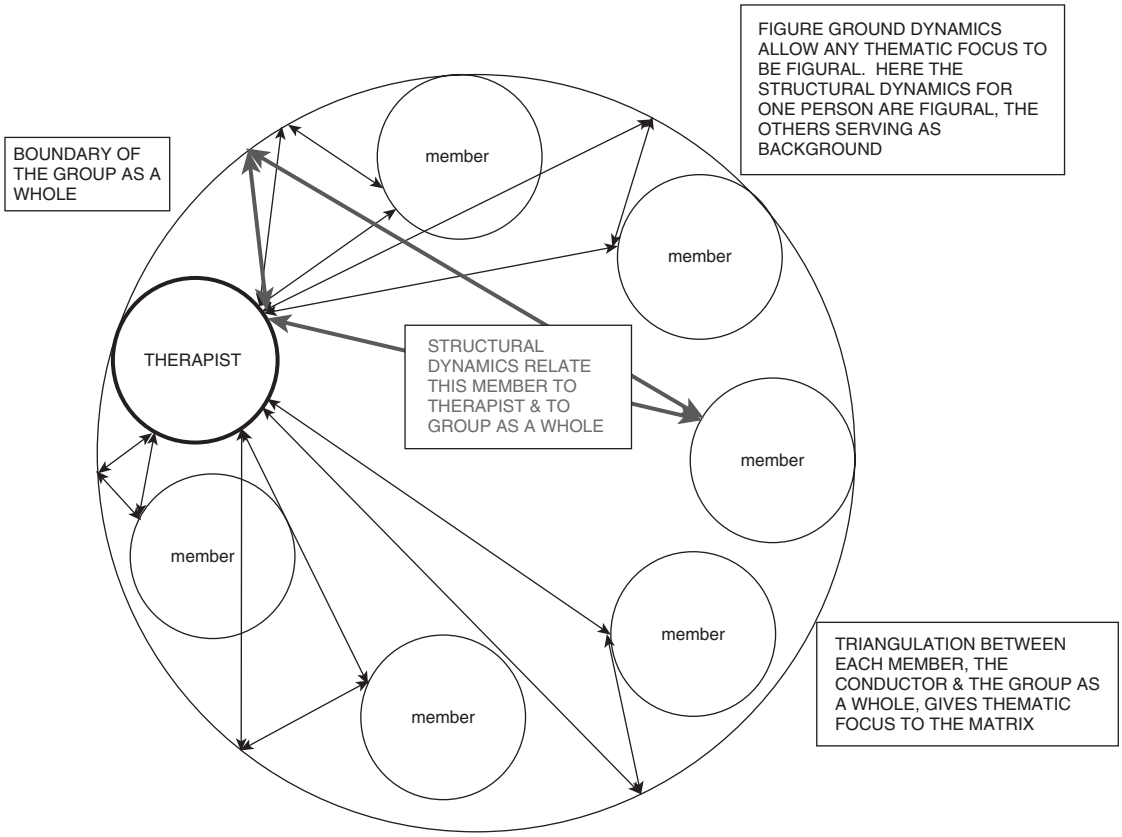


Fig. 6.3.6.6 The elements of a group-analytic theory.

in which it takes place. Figure 6.3.6.7 illustrates the conductor's therapeutic role in relation to each of these dynamic elements of the discussion.

(f) Recent developments in group-analytic psychotherapy

The role of dreams in group psychotherapy has been reconsidered.[48] Friedman described two of the unconscious functions they serve in group therapy—requests for containment and influence on relations with the dream audience.[49] Lipgar and Pines[50] work towards an integration of the ideas and methods of Bion and Foulkes with discussion from an international panel. Nitsun[51] considers sexuality in group psychotherapy: sexual identity, boundary transgression, erotic connection, dissociation of desire, the group as witness, erotic transference and counter-transference, and the effectiveness of psychotherapy.

(g) The conductor as therapist and group member

At times of coherence, when the members are close in the shared experience of a moment, or when there is an issue charged with meaning, the group-analytic approach comes into its own. The symbolic content of the discourse might evolve in the language content, the flux of interactions, or the attention given to an individual's problems. The conductor needs to be able to model this use of imaginative play—with images, associations, or exchanges—and then stand back to allow members to take the enquiry forward. Cox[52] shows how images can safely hold experience too painful or brittle to tolerate much analysis. People discover that images can touch the depths before they stir the surface, giving access to profoundly felt and deeply hidden concerns. When used and played with in this way, **the mutative use of metaphor** provides the whole group with a vehicle for change. Other aspects of the therapist's activity are summarized in Table 6.3.6.4.

The dynamic life of groups

In the sections that follow we draw on our own Group-Analytic model to examine a range of clinical considerations and make them as relevant as possible to the widest range of practitioners, regardless of their own models.

Group development theory

Many schemata have been proposed, usually derived from time-limited experiential and study groups but some of these can provide a useful orientation. Bennis and Shepard,[53] followed by Yalom, write about the initial stages of **orientation**, involving a search for structure and goals, dependence on the leader, and concern with boundaries. Their second stage is characterized by **conflict**, particularly over **norms**, authority, and control. Their third stage is achieved with a high level of group **coherence** that allows for inter- and intra-personal exploration. They acknowledge that the boundaries between phases are not clear and that a group never graduates permanently from any one phase.

In slow open groups, especially, this idea of discrete phases has its limits. The group-analytic approach provides a cyclic model in which focal issues such as affection, intimacy, personal history, and change, are understood and struggled with repeatedly. A group

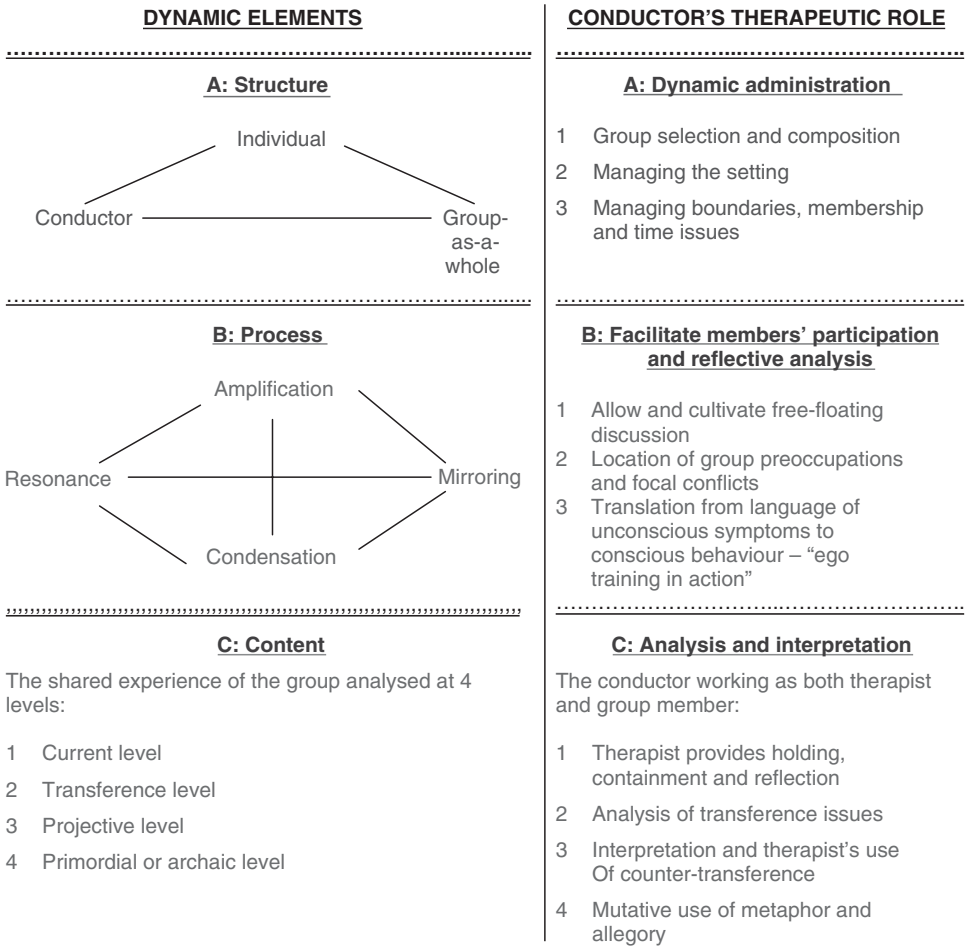


Fig. 6.3.6.7 Structure, process, and content—the conductor's therapeutic map.

Table 6.3.6.4 The therapist's activity

A: Leadership and analysis
1 Model a capacity for open, direct communication
2 Maintain therapeutic neutrality
3 Attend to boundary events
4 Provide holding and containment
5 Withhold personal material
6 Drawn on counter-transference for
7 Reflection on group events
8 Bring events from background to foreground, or vice versa
9 Provide linking communications
10 Clarification and confrontation with individuals
11 Attention to omissions, avoidance, denial
12 Maintain silence
B: Interpretation
1 Locate group preoccupations
2 Translate from the language of unconscious (individual and group) behaviour
3 Interpret or provide metaphorical constructions for
I Defences and resistances
II Transference and projective process
III Archaic and primordial experience

understands in these terms struggles with **developmental tasks** rather than phases, in the course of which it is the individuals who enjoy growth, differentiation, and progressive change.

(a) Developmental stages and thematic focus

Figure 6.3.6.8 provides a map of these developmental tasks conceived of in logical rather than sequential terms. The way in which one person comes into the group will be different to the arrival of another. The terms on which one person joins will have a determining influence on each of the stages they pass through. For example, preoccupations about how someone came into the group—the terms of their engagement—might be resolved only when the person leaves perhaps one, three, or five years later, when they have to assess the outlook for their future as they look back over the years spent in therapy. So the diagram can be used as a thematic map for one person's journey through therapy, or as a way of appraising the stage reached by the group as a whole.

Early stages are dominated by the anxiety of being involved in a new situation and by questions about other group members. Preoccupations are likely about boundary issues, confidentiality and security. In the second stage, those familiar with psychotherapy—or otherwise accustomed to talking about themselves—will be at an advantage. There will be a range of tensions about group norms and disclosures, reasons for joining and discrepant levels of confidence about using the group. In the third stage, members will defend against intimacy with one another and struggle with questions of trust, attachment, and affiliations. In the

fourth stage, as members become increasingly able to trust the group with self-disclosure, observed changes might become manifest as self-exploration yields the beneficial experience of individuation and differentiation. In the concluding stage, people prepare for departure and find themselves comparing points of difference between changes achieved in the group and the state of their lives outside—generalising from the arena of therapy to that of real life. Has therapy made a lasting difference? Will it be maintained outside?

The language of the group

Foulkes suggested that 'symptoms, in themselves unsuitable for sharing, exert, for this very reason, an increasing pressure upon the individual to express them'.[54] The group equips the person to transform the mute and inchoate language of symptoms into a socially understandable form of discourse. Following publication of Schlapobersky's paper, The Language of the Group, there is increasing interest in characterizing group phases in terms of the language that predominates, using theory from discourse analysis and the Foulksian concept of free-floating discussion.[16]

It is possible to differentiate between three primary forms of speech that arise in the matrix of any group. At the most basic level **monologue**—speaking alone (with or without an audience)—is a form of individual self-expression. At the next level **dialogue**—a conversation between two people—is the form of communication that distinguishes a bipersonal exchange. And at the third level **discourse**—the speech pattern of three or more people—allows the free interaction of all its participants in a flexible and complex exchange that distinguishes the communication of a group. These patterns of speech are universal cultural forms arising in all communication and are present in the life of every group, although in no set order. Monologue can be understood as a soliloquy, dialogue as the resolution of opposites or the search for intimacy, and discourse as the work of a chorus. The use of free-floating discussion allows a pattern of exchange to move freely between these different speech forms, each of which constitutes a distinctive type of communication. It is through this movement—from monologue through dialogue to discourse and back again—that the group-analytic method comes into its own, creating an arena in which the dialectic between the psyche and the social world helps to refashion both.[13,14]

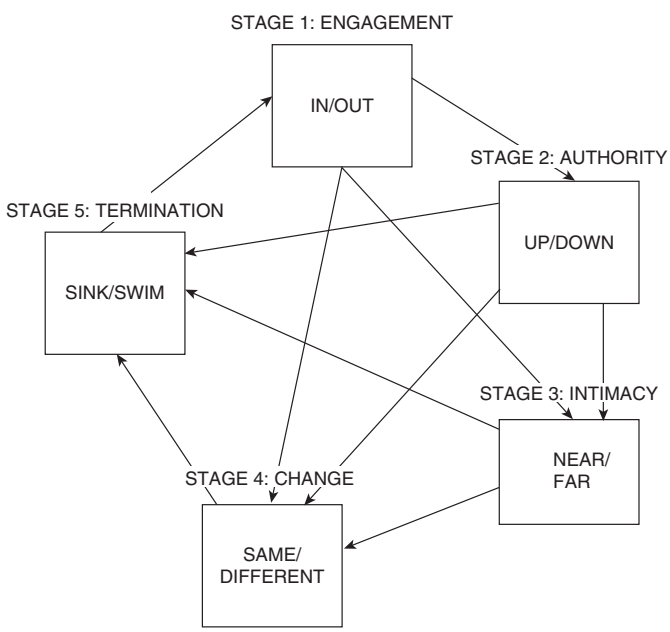


Fig. 6.3.6.8 Developmental stages and thematic focus.

Leadership

The group analyst works as both group member and therapist, beginning with dynamic administration, assuming an active role in a new group and allowing a **decrescendo** of his own role as the group gains authority. He is responsible for helping the group with the **location** of disturbance in its process and for providing a balance between analytic and integrative forces whilst manifest content is translated into language that describes the unconscious. Transference is prominent but the work is undertaken in the dynamic present. Foulkes' account in the following passage,[55] of the conductor at work in a group, stands in dramatic contrast to Freud's account of the psychoanalyst at work behind the couch:

> He treats the group as adults on an equal level to his own and exerts an important influence by his own example . . . representing and promoting reality, reason, tolerance, understanding, insight, catharsis, independence, frankness, and an open mind for new experiences. This happens by way of a living, corrective emotional experience.

Disturbed, narcissistic, or borderline patients bring to the group more primitive psychic structures and processes that put strain on the resources of other group members. Such patients can create turmoil in which the leader's task is to maintain the group at a more mature level of psychic organization. By responding to part-object relationships and processes on the level of whole-object relations, containing responses can be established. Progressively, these help to build for the disturbed patient a more benign world of inner object relationships and processes. More disturbed patients desperately seek attention in ways that are inappropriate and disruptive. This search for **attention** arises because the patient cannot establish a sense of **connection** between herself and the processes of the group. Mirroring and resonance can steadily come to replace these isolated and fragmentary responses allowing the patient to attain—for the first time—a coherent sense of self and a capacity to recognize the identity of others.

(a) The dynamics of change

As noted above, Yalom cited 11 'curative' factors responsible for change in groups. Foulkes believed there were four key group-specific factors: mirroring, exchange, social integration, and activation of the collective unconscious. Tschuschke and Dies[56] have identified five key factors that they could study empirically (Table 6.3.6.5).

(b) Binding forces: alliance, cohesion, and coherence

Group alliance has as its major focus the quality of the relationship that develops between each individual member and the therapist(s). These alliances grow as each individual develops a transference relationship to the therapist and can feel that their own particular individual dynamics are recognized. Group cohesion used to be compared to the concept of therapeutic alliance in individual psychotherapy but current research shows that alliance (defined as the affective bond that develops between each group member and the therapist) can be differentiated from group cohesion.

Group cohesion describes the bonds between group members, their attitudes, and their commitment to therapeutic work, in particular

Table 6.3.6.5 Five factors in the dynamics of change open to empirical evaluation[56]

1 Cohesiveness	Closely related to 2 & 3 as determining factors that influence outcome in 4 and 5
2 Self-disclosure	Closely related to 1 & 3 as determining factors that influence outcome in 4 and 5
3 Feedback	Closely related to 1 & 2 as determining factors that influence outcome in 4 and 5
4 Interpersonal learning output	Evidence of working through process seen in 4
5 Family re-enactment	Findings of change in these patterns provides evidence of enduring psychic change

their feelings of attraction and dependency. These processes are interdependent and in combination they provide optimal conditions for positive group process and outcome.[57]

Group coherence is a more evolved group state requiring but going beyond cohesion. It evolves as a semantic matrix, built on the earlier, relational matrix. When a group moves through support and understanding to be able to recognize and work through conflict, it can achieve a sense of containment.[58] At this stage the group becomes a more complex, self-evolving, and self-defining entity capable of reaching deeper levels of exploration, acceptance, and understanding.

(c) Corrective emotional experience

The concept was introduced by Alexander to describe the patient's recognition of discrepancies between present and past experience. Grotjahn, a colleague of his, applied this to the group, describing it as 'the corrective family experience'.

> There is a built-in correction of the transference phenomena through the peer relationship in groups. An analyst is trained to let the transference neurosis to grow to full bloom. Members of a group are neither trained nor willing to accept such projections . . . and will correct them. This is the basis of a corrective therapeutic family experience.[59]

Oedipal, sibling, and pre-oedipal constellations are activated in group therapy and can be worked through as the members play parts in each other's family scenarios.[60] The task of the therapist interprets transference and helps the group as a whole to develop coherent norms of understanding and responsiveness that equip its members to go beyond the private preoccupations that might have brought them into therapy. Garland calls this **taking the non-problem seriously**.[61]

(d) Resonance, mirroring, and other dynamic processes

Foulkes described the group as 'a hall of mirrors'. Each person, he thought, could see aspects of themselves reflected in the personality and behaviour of others and could often more easily recognize these aspects of the self than by direct introspection. We cannot know ourselves in the absence of reflective mirroring, the feedback and information we obtain from others according to our presence and behaviour in the group brings us to a greater awareness of who we are. Resonance in the group is the unconscious communication of emotion by which the other process dynamics—like mirroring, amplification and condensation—are effected.

Insight and outsight, regressive and progressive forces in groups

The dictum, 'where id was, ego shall be', coined originally by Freud to ground the psychoanalytic enterprise in an easily comprehensible principle, has been superseded for group workers by the idea of **ego training in action**, coined by Foulkes. It emphasizes two related issues. The first is that **insight**—the understanding of the self—is related to **outsight** defined as the understanding of the other(s). And the second is that people find growth and change as much in what they do for others—**progressive behaviour**—as in what others can do for them—**regressive behaviour**. So the group is designed to create an arena in which there is continuous flux between these different elements.

(e) Imitation, identification, internalization, and differentiation

(i) Imitation

From early in infancy, we observe and copy what others do. A therapeutic group is designed to create conditions in which imitative behaviour, or modelling, can be experienced, monitored, and understood. It brings the group members closer together and, as it develops into identification and internalization, it increases the cohesiveness of the group.

(ii) Identification

Peoples' predominant modes of relatedness to one another—such as compliance, avoidance, dominance, receptivity, exploitation, and need—become apparent in the group. The group exposes these patterns of relatedness to reveal how the internal objects with which members have identified can be externalized and encountered in the group through projection and introjection. Members compel each other to play and re-play the dramas of their interior lives and past injuries in the 'here and now' of group experience.

(iii) Internalization

Intimacy with others is developed through the exchange of understanding. As it is given and received, it allows for new forms of intimacy within the self. **Constructive** experience in the group is taken in and becomes part of the self. This can allow the recognition and reappraisal or **deconstruction** of repetitive patterns and fixed, maladaptive characteristics. And this can lead to the **reconstruction** of the self in the concluding phases.

(iv) Differentiation

Over time people become aware of significant differences in their reactions, emotions, and psychological structures. Examples are increased tolerance of affect, understanding and modification of self-inflicted pain, diminution of guilt and shame, retrieval of lost aspects of the self, increased openness with others, increased spontaneity and creativity, and, most critically, feelings of tolerance for or forgiveness of the self, for the constraints into which it has been driven by past injury and present defence.[62]

(f) Personal and group resistances

Some resistances are manifestations in the group of the characteristic defences of the members, evident in their interpersonal behaviour. Others are resistances of the group as a whole, shown in blockage of free-floating discussion, opposition to group interaction, sub-grouping, and opposition to the deepening and broadening of the group's exploration. A constructive function of resistances should be kept in mind when monitoring the pace at which both individuals and the group can progress without experiencing

overwhelming anxiety. Resistances can protect people from fears of loss of self and identity; and from fears of engulfment through excessive intimacy. Resistances then become opportunities for understanding fear of change, for the re-working early developmental patterns, and for discovering the freedom that can follow release from excessive internal control.

(g) The anti-group

Nitsun developed an understanding of the anti-group, bringing together the work of Bion and Foulkes.[63] He introduced the concept to help understand the negative experiences therapists face in periods of stagnation, hostile silence, severe conflict in or premature departure from the group, or negative feelings about the work of the group. These situations can arise when the group recapitulates early experiences of loss, deprivation, anger and envy or in an effort to avoid these emotions. Negative emotions can then be projected towards the group and the conductor who represent early care-givers. These feelings can be worked on when the therapist uses counter-transference to recognize and verbalize projected feelings. The idea of **negative elaboration** (see below) is a useful guide in this process.

Basic clinical issues

Dynamic administration

(a) Selection and composition

Group composition is the therapist's first and most enduring contribution to the group for its membership will determine the outcome of therapy. Preparing patients for treatment with several individual sessions or, if necessary, an extended programme of preparatory work will provide the therapist and patient with a basis for judgement about therapeutic prospects. Preparatory work has been found likely to enhance participation and reduce drop-out rates, although findings are not conclusive.[64]

The criteria for selection are exclusive rather than inclusive since most patients seeking psychotherapy can be accommodated in a group, provided a suitable one is available. The selection process should take into account both the patient needs and the composition of the group. A service should, ideally, provide a selection of groups into which people can be placed both according to their needs and characteristics, and those of the particular group. Selection criteria aim to optimize the 'fit' between the needs and resources of the individual and those of the group.

Table 6.3.6.6 lists inclusion criteria, while Table 6.3.6.7 gives a shortlist of excluding criteria. Lists of this kind should be used with caution. In general, at least four inclusive criteria should be found amongst those to be included in outpatient, dynamic psychotherapy groups with a mixed population. If there are four exclusion criteria, one should be very wary about including the person in a group. However, there are many exceptions.

The criteria in Tables 6.3.6.6 and 6.3.6.7 hold good for mixed groups and outpatient services generally. With homogeneous groups for special populations, the range of potential candidates is much wider, for example, Hearst describes a population of severely deprived mothers drawn entirely from those on the exclusion list.[65]

(b) Homogeneous and heterogeneous groups

Homogeneous groups: Are for people with similar symptomatic or diagnostic pictures, such as phobias, anxiety, or depression. Such groups offer more immediate support to members, are better attended, and provide faster symptomatic relief. However they may remain at a more superficial level with less interpersonal learning.[67]

Homogeneous groups are also used for those with similar personality structure or life-history, particularly those in socially extreme categories. For example, men with histories of sexual violence or women who have suffered rape or torture may be treated.

Heterogeneous groups: Melnick and Woods[67] suggest that group composition should be guided by an optimal balance between conditions ensuring group maintenance or homogeneity, and those maximizing interpersonal learning or heterogeneity. A group which shares one strong characteristic—a diagnosis such as an eating disorder, or a personal attribute like intelligence—can accommodate a diversity of presenting problems or social backgrounds. If, on the other hand, the members are similar in social background, diversity can be incorporated on another basis, such as diagnosis. Thygesen[68] found that diversity enabled group members to recognize and work with differences in mental and emotional attitudes, life histories, and developmental problems. Recognizing and working with difference develops emotional resources, promotes flexibility, and the tolerance of emotional tension. And it encourages the group to move from **cohesion**, in which security is based on identification, to **coherence** in which relationships are based on differentiation.

It is useful also to identify members likely to be isolated from the rest of the group by age, ethnicity, gender, personality, or problems or a history that noone else shares, for they are likely to find

Table 6.3.6.6 Inclusion criteria for psychodynamic groups (at least four of these criteria should be present, for someone to join a group)

1	Motivation to address personal issues, to resolve problems
2	Willingness to try and participate
3	Some experience of successful relationships in childhood or present
4	Some interest in exploring and understanding the self
5	Some capacity to talk, listen, and relate
7	Some interest in others
8	Some sense that being amongst others could be helpful
9	Some ability to sympathize or empathize with others' needs and problems
10	Some indication of future reliability in attendance

Table 6.3.6.7 Exclusion criteria for psychodynamic groups[66] (if four or more of these criteria are present, then questions should be raised about inclusion)

1	Those in acute crisis
2	Prior history of broken attendance in therapy
3	Major problems of self-disclosure
4	Major problems with reality testing, i.e. paranoid projections or psychosis
5	Pathological narcissism
6	Difficulties with intimacy generalized into personal distrust
7	Defences that rely excessively on denial and disassociation
8	Emotional unavailability
9	Tendency to be verbally subdued or withdrawn
10	Tendency to be hostile and aggressive, verbally or otherwise

the group experience threatening. We do not put a patient into a position of being isolated.

Managing the structure, setting, and time boundaries

(a) Optimal and sub-optimal size

The optimal number for small group psychotherapy, ranging from five to nine, is determined by practical considerations. A smaller number than this minimum is likely to have an active attendance of only two or three and will not necessarily generate the corporate energy to produce movement. A group larger than this will exceed the number that can be taken into one person's confidence in a face-to-face exchange.

Sub-optimal groups can provide valuable therapy under good conditions.[69] Low and irregular attendance is often associated with a problematic composition, particularly if there is a high proportion of members with character disorders and borderline features. Their inner sense of deprivation and loss can make the group seem unreliable and threatening which can be reinforced if the therapist is seen to be in difficulty. Therapists who can maintain an understanding attitude and positive commitment to the group's future will usually find the situation settles into a working nucleus of members that can then be built upon.

(b) Setting and time-boundaries

The therapist supplies, creates, and maintains the setting throughout the group's life. This requires attention to such matters as meeting times, punctual beginnings and endings, confidentiality, the predictable frequency of its meetings and breaks, and the general guarantee of a stable background.

Every aspect of the group's life, including absences, departures, late attendance, and extra-group communication in terms of letters, phone calls, and messages, referred to generically as **boundary events**, are open for discussion. Boundary events are interpreted for the meaning they might hold for the life of the group as a whole. This is a task initiated by but not confined to the therapist. Also, the setting itself—to which each member is seen to contribute—comes to acquire a capacity to **hold** the individual members and **contain** their anxieties and insecurities. The issue of containment is of great importance, as is the discovery—by people who may have no belief in themselves as responsible members—that they can take responsibility for themselves and expect responsibility from others. The group is thus responsible not only for the nurture, acceptance, and security of its members, but also for their containment, the setting of limits and the maintenance of consistent authority. The therapist will often have to lead the way in modelling both roles for the group's members.

Fostering therapeutic norms and a culture of enquiry

The therapist encourages the sharing of experience and helps to balance participation, recognition, and translation. The thrust of a group's life is towards greater shared involvement and the expression of emotions. The expression of feeling may arise in the recounting of members' life-situations and their reasons for therapy—**narrative emotion**—or it may arise in the interpersonal encounters engendered by the telling—in the **drama** of 'here and now'. Therapy becomes effective when the problems that brought the patient into treatment become recognizable in their interpersonal encounters. At this point the affect lodged in the narrative will interact with the drama of the group's current emotions, creating opportunities for corrective emotional experience. As the group progresses from **constructive** to **deconstructive**, and ultimately to **reconstructive**, experience, it will encompass gesture, behaviour, body-language, and other non-verbal communication, and actions that convey feelings when emotions have no words.

Guidelines for intervention

There are four modalities of time and place in any therapy group. The content of the exchange might be located in the past outside the group, the past inside the group, the present outside the group, and the present in the group.

Free-floating discussion will carry the focus of the exchange between these different modalities. The therapist's task is to follow the interaction, to use interventions sparingly and strategically, to cultivate a reflective curiosity for which Table 6.3.6.8 offers some pointers, and to work towards a progressive shift in the focus of attention, from no. 1, in Fig. 6.3.6.9 towards 4, via 2 and 3. A group governed by narrative in its free-floating discussion is likely to be dominated either by the the group's own past or by the past of its members outside the group or by their current lives outside. A group governed by members' intense experience of one another in the present, is likely to be governed by the drama of immediate encounter and understanding. The corrective recapitulation of early family life is likely to arise when the issues that predominate in no's 1, 2, and 3, are translated into issues that prevail in 4, where they can be addressed in the present which allows new resolutions to be forged.

Groups for special populations

Nine distinctive clinical populations or approaches merit special mention. Some pose distinctive problems for group workers and, as we indicate in the section on borderline patients, they sometimes need co-therapists working together in the group.

Table 6.3.6.8 Interventions in time and place

1	Address process rather than content
2	Help members recognize aspects of themselves in others and accept the viewpoints of others on themselves
3	Monitor the intensity of participation to allow an enabling pace so members can develop resources to deal with the intimacies of each others' lives
4	Establish a sense of enquiry about the fluctuations of mood and outlook, and a sense of curiosity about thematic movement between the different modalities of time and place
5	Help the group recognize role configurations taken up by individuals (therapist's assistant or rival, joker, complainer)
6	Hold in mind the gestalt of figure and ground. If someone stands out, what is the background against which they do so? If the ground changes the underlying pattern of the group, how might the figural person be affected?
7	Help the group recognize sequences and patterns
8	Help the group decode and find meaning in the constructive, deconstructive, and reconstructive elements of its exchange
9	Work towards the coherency of the group
10	Foster the integration and integrity of its members

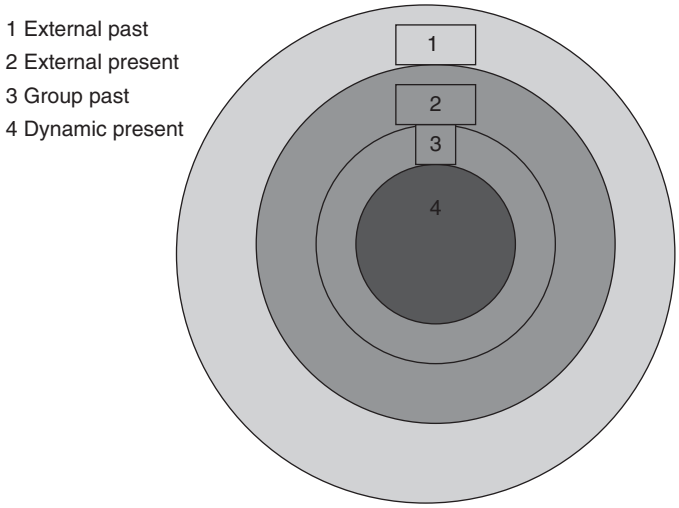


Fig. 6.3.6.9 Location of group discourse in time and place.

Intensive outpatient psychotherapy heterogeneous groups for mixed populations

Group analytic psychotherapy has come into its own as a treatment modality in out-patient settings in the context of clinical agencies in public and private Health Services. Table 6.3.6.9 gives a picture of patients' clinical needs in the private practice of John Schlapobersky over a twenty-year period. Some attempt has been made to indicate how these categories are mapped against those in ICD 10 and DSM IV.

The medically ill

Groups have been used successfully for people with diseases of the cardiovascular, endocrine, gastro-intestinal, pulmonary, neurological, and renal systems. Amongst the applications are groups for those with renal dialysis and organ transplantation; chronic conditions such as diabetes, irritable-bowel syndrome, and rheumatic disease; and life-threatening conditions such as metastatic cancers.

Groups usually have members with a single disorder, last for 20 sessions or fewer, and are psycho-educational in purpose. They emphasize compliance with treatment regimes and opportunities to share and explore the affective consequences of the conditions. The focus is on helping patients come to terms with the complaint; make changes in lifestyle, including diet, exercise, work, and recreation; deal with the inevitable anxieties about death, stigma, demoralisation, feared dependency, and loss of control; and resolve issues in relationships with loved ones and care staff. Clinicians write of how, despite their brevity, these groups provide a forum for sometimes profound exchange about major existential issues including pain, fear, and impending death.[70]

Evaluation indicates such groups are effective in improving quality of life and, to some extent, in prolonging its duration even with metastatic cancer.[71] Improved survival correlates significantly with enhanced, active coping. Spiegel reported a significant, near-doubling of survival effects in 86 women with metastatic breast cancer treated for one year with weekly, supportive-expressive group therapy.[72,73]

The chronic mentally ill

In North American literature there are consistent findings that schizophrenic patients do best in support groups that help them manage their symptoms and devise coping strategies for day-to-day problems.[74] Group discussion is used to help overcome problems of social isolation, develop social skills and coping resources, and learn about maladaptive interactions evident in the 'here and now' of the group. The leader creates a sense of safety and security with supportive feedback. Areas that people find helpful include the management of symptoms, expression of emotion, and relationships with others. Groups are less valued for acquiring insight or receiving guidance about the illness, medication, or economic problems. Interpretations that reveal or explore unconscious conflicts, particularly transferential issues between members or with the therapist, are more likely to be harmful than helpful.[75] Patients respond positively to an interactive, open, and safe group environment, with gradually increased cohesion, decreased avoidance, and decreased conflict.[10]

Recently, Resnik[76] has described the use of insight-orientated psychotherapy with delusional problems using psychoanalytic principles in long-term group psychotherapy for psychotics.

Borderline personality disorders

Working with these patients can pose difficult clinical problems based on their 'stable instability', including their inability to form a stable therapeutic alliance, mood oscillations, poor impulse control, limited reflective ability, painful states of psychic emptiness, followed by self destructive acting-out including self-destructive gestures and self-mutilation, violence; substance abuse, sexual acting out, and eating disorders. Some attempts at short-term group therapy combining cognitive behavioural approaches with the psychodynamic have shown promise. Longer therapies in out-patient groups have been encouraging but have not been evaluated. Good results are reported with in-patient applications at, for example, The Cassel and Henderson Hospitals, and in other forensic settings where day hospital group treatments are followed by outpatient group therapy in a carefully monitored programme.[77]

Homogeneous outpatient groups do not do well because the members use similar 'primitive' defences of splitting, projection and denial, and have not acquired a stable sense of identity. They are suspicious of close interpersonal contact and have little capacity to care for or be cared for by others. Such groups lack cohesion, are pervaded by a sense of pessimism, hostility, rivalry, and easily fragment. Self-destructiveness is turned onto the group and its therapists who may experience intense counter-transference feelings of frustration, despair, and anger.

These problems can be managed in the containing setting of closed or specialized institutions. However, it is generally better to place one or two borderline patients in an otherwise well-functioning group where other members will not always respond to borderline pathology at the same primitive level. They may—like the adult carers of children—respond with understanding, find ways of setting limits, and expect co-operation with the task at hand. After sometime, lengthy periods in which borderline patients maintain a frustrating presence on the margins of the group, or make themselves felt in aggravating terms at its very centre, they can acquire resources to take part in the group's work in more mature terms. This combination of borderline and neurotic members in a carefully composed group can benefit both parties. Borderline members will often have an unerring accuracy of perception about others, they can shake the group into more active interactions, and may not collude with others' neurotic defences.

Table 6.3.6.9 Categories of clinical need catered for by intensive, outpatient group therapy with mixed populations (many of those who join groups are found in more than one of these categories of need)

1. High dependency need including:	2. Problematic reactions to traumatic life events including:	3. Character problems including:	4. Serious relational problems including:	5. Selected people with borderline personality disorders including	6. Training of mental health professionals including:	7. Those seeking personal growth or understanding including:
a Recovery from serious or enduring mental and physical illness b Help with the resolution of moderate psychiatric symptoms c Resolving addiction and substance abuse	a Loss, injury, illness, disorder, infertility b Massive psychic and physical trauma c Adult sequelae of child sexual abuse	a Immaturity and developmental problems b Chaotic life situations c Problems of identity and meaning d Problems of gender, sexuality, and orientation e Occupational problems	a Intimacy avoidance b Recurrent broken relationships c Intractable conflicts	Those with insight who both need and can tolerate containment	Psychotherapists, psychiatrists, psychologists, social workers, clergy	People who would have previously sought this through individual psychoanalysis

(John Schlapobersky: Group-Analytic Practice & Southwood Practice: 1987–2007)

Category 1: Corresponds to Axis I in DSM IV and Groups F1–4 in ICD 10.

Category 2: Corresponds to Axes III, IV, & V in DSM IV and Group F5 in ICD 10.

Category 3: Corresponds to Axes IV and V in DSM IV and Group F8 in ICD 10.

Category 4: Corresponds to Axis V in DSM IV and Group F8 in ICD 10.

Category 5: Corresponds to Axis II in DSM IV and Group F6 in ICD 10.

The containing resources of the other members that set the norms and values of the group can slowly be internalized by the borderlines. As with many of these specialist areas, therapists working with these populations will need to acquaint themselves with the literature, have ongoing supervision and may—to begin with—work more fruitfully in co-therapy.

For a review of group psychotherapy for personality disorders see [78, 79].

Forensic groups

Group therapy has been used for forensic populations, previously thought untreatable. Pioneering work has been done in the UK in special hospitals,[80,81] in prisons such as Grendon,[82] and great strides have been made in outpatient services at the Portman Clinic in London.[83] Therapeutic community treatment for mentally disordered offenders in North America and Western Europe has been reviewed by Lees, Manning, and Rawlings.[84]

The aim of forensic group therapy is to help patients find words, rather than actions, to express impulses and compulsions. A major part of the group's work is to provide psychic space for perspective, negotiation, recognition, acceptance, and verbalization of hurt. Sharing the past and present can make it accessible as internalized, persecutory, and vengeful monologues are brought into dialogue.

In Britain, the Probation Service is the agency most actively involved in the groupwork with offenders.[85] A report for the NHS Centre for Reviews and Dissemination (CRD Report 17) reviews studies of therapeutic communities and small groups and indicates that these show the most promising results of any form of treatment for anti-social personality disorders.[86]

Trauma

Group work with trauma victims is a comparatively new field but one in which there is a vigorous range of applications. The most comprehensive overview is Van der Kolk's text.[87] There are groups for survivors of sexual abuse,[88] war trauma,[89] and torture and other forms of organized violence. Group work applications to traumatized children and adolescents is discussed in reference 90 and more generally in reference 91.

At the Medical Foundation for The Care of Victims of Torture in London, a group work programme caters for refugees and asylum-seekers. The groups provide psychotherapy for massive psychic and physical trauma, for problems of displacement and exile, and for trans-cultural problems.[92] The diversity of different kinds of groups, including a range of activity, problem-solving, and psychodynamic groups, ensures that people can be provided with an environment in which they can each realize their own potential for self-healing. The programme is grounded in commitment to human rights, to team-work that addresses counter-transference issues, and to principles of positive intervention to counter emotions of hopelessness, and despair.[93,94]

Couples groups

These groups cater for people in stable but troubled relationships in which there is some form of pernicious collusion. The approach can provide symptom relief and personality change in even severe difficulties with relationships that last but do not work. Family therapy is concerned with the systemic function served by symptoms. Psychoanalysis is concerned with the origins of the

symptoms in object relationships. Group analysis provides a bridge between these paradigms, helping therapists find a point of intervention between marital and object relationships.

In a group, the process moves between the psychology of the individuals and the dynamics of their marriages. This interplay between the pair and the person—the interactive and historical dynamics—is part of the group's free-floating discussion. The therapist uses this interplay, following the patterns of the group's progress and making interpretations. Work is at times transferential, other times, systemic techniques are used to address immediate issues.[95] Themes to which the group resonates are amplified and members are helped to condense from this discourse, the kind of personal knowledge that promotes growth and change. There is insufficient evaluation of the method's efficacy, but experience in hospital and private practice is very encouraging.[96]

The elderly

This is an area of relatively new therapeutic exploration. A range of groups has been found useful in helping the aged to face their problems, improve their functioning, and feel happier about themselves. The groups include those offering support, activity, psycho-education, problem-solving, and insight, and takes place in many settings. Common themes include the ubiquity of loss, the acceptance of death, and the value of humour in lightening mood.[97]

Brief therapy in groups

Spurred by economic and managerial pressures a distinctive modality has evolved. Sessions are held weekly and number between 6 and 30. Characteristics that differentiate brief from long-term therapy include: clearly defined therapeutic goals agreed at the outset, the early establishment of a therapeutic alliance, active and flexible therapeutic style, a focus on 'here and now' group process, the maintenance of time awareness, monitored in stage-specific terms, and the vigorous, directed exploration of thematic content.

Short-term groups may be homo- or heterogeneous. Homogenous groups have proved effective in helping patients deal with loss and grief, the consequences of trauma and abuse, and common problems coping with physical illness and disability. Heteregenous groups require more psychodynamic commonalities such as shared problems in inter-personal relationships, the ability to recognize and work on psychological issues, and the ability to cope with the speed and intensity of the process. Those who lack psychological sophistication or are not motivated for self-exploration, are not suited.

Research and evaluation

The general trends in literature show ample empirical confirmation that group treatments represent a powerful therapeutic intervention. A comprehensive overview of literature described positive outcomes with alcoholism, anxiety disorders, bereavement, bulimia, depression, schizophrenia, and sexual abuse.[98] There is also evidence of the adverse outcomes in group psychotherapy which can guide the clinicians training for the work.[99]

(a) Outcome research

A meta-analysis of 58 controlled studies of psychotherapy for the treatment of depression showed that in comparison to a waiting list control the average treated patient was better off than 80 per cent of the controls.[100] The efficacy of group and individual therapy was almost identical. Tyllitski[101] also reported no appreciable difference between individual and group therapy effectiveness, both doing better than the control condition. Budman et al. found significant improvement in time-limited individual and group therapies. It seems that most patients who are suitable for psychotherapy will benefit in either modality.

(b) Process research

(i) Preparation of patients for group therapy

In a review of 20 controlled or comparative studies, Piper and Perrault[102] found that preparation has a positive effect upon attendance but that it could not be shown to have a direct effect on outcome.

(ii) Therapist activity

Development of constructive group norms will depend on factors such as careful group composition and leadership style. Foulkes' idea that the conductor has, at first, a relatively high rate of activity which decreases as the group develops its own resources for psychological work has been supported by later research findings.[103]

(iii) Group process variables

Research on therapeutic factors (e.g. self-disclosure and feedback) and leadership technique indicate that members in well-established groups are engaged in many different types of psychological work. They are less group-centred and more likely to be confronting the personal distress and maladaptive interpersonal styles that brought them to treatment in the first place.[104]

(iv) Therapeutic factors

In a study of long-term in-patient groups, Tschuschke and Dies[105] investigated five therapeutic factors: cohesiveness, self-disclosure, feedback, interpersonal learning-output, and family re-enactment. All five therapeutic factors were associated with clinical improvement with group cohesiveness, an important ingredient. They suggested that **affective integration** into the group, that is the high and positive emotional relatedness to co-members, promotes the capacity to disclose and leads to more frequent and intense feedback from fellow patients. It appeared that feedback given earlier in the group had a stronger relationship to treatment outcome. This may suggest that **interpersonal feedback** needs time to be assimilated and worked through before it can be utilized effectively. There are significant differences between successful and unsuccessful patients in terms of level of group cohesion and amount of self-disclosure. Patients who disclose little and do not feel drawn to the group receive relatively little meaningful interpersonal feedback and become neglected. They concluded that **cohesiveness, self-disclosure**, and **feedback** and together promote **interpersonal learning** within the group.

These findings were confirmed by the author's later studies, the most recent of which was published in 2007[106] and by recent independent studies using different parameters; see, for example, references [107–109].

Conclusion: planning a service

For group psychotherapy to be effective the group has to be the primary focus of therapy; patients need to be well selected; and

therapists need to be adequately trained. Therapeutic competence is not a function of mastering the literature so much as it is the outcome of experience in the group situation itself. Courses introducing different group methods are now widely available throughout the UK, Continental Europe, and the USA. Group training is also provided in many general training programmes and, along with clinical supervision, is offered in many health and other service agencies.

Long-term **outpatient** group therapy of 100 sessions or more is effective and economic in producing lasting benefits for patients with a wide range of medical and psychiatric symptoms, interpersonal problems, traumatic life experiences, character and personality disorders. **In-patient** group therapy is an effective resource in the context of acute units working with crisis, and in secure units working with long-term problems. **Short-term** group therapy for selected conditions requires careful composition of the group and an active, flexible therapeutic approach.

We have not tried to cover the range of group services and approaches for children. Chapters by Schamess[110] and by Kymissis[111] cover group work with children and adolescents, respectively. Further reading is available in Evan's text[112] and Melzack.[113]

There are economic arguments for group therapy. In one study, the quality of improvement between individual and group therapy of psychiatric patients was not significantly different but the cost of the service was different. Cost savings were calculated as the reduction in medical consultations and hospital attendance, and lost workdays. For those treated with psychotherapy of any kind, the cost of treatment was 25 per cent per patient less than it was for those who did not receive psychotherapy. The cost of group psychotherapy per patient was about a third less than for individual psychotherapy.[114]

Further information

A. Bateman, D. Brown and J. Pedder's (2000) *Introduction To Psychotherapy* (Routledge) sets group therapy in the context of other psychotherapies. D. Stock Whittaker's (1995) introduction, *Using Groups To Help People* (Routledge) and M. Aveline and W. Dryden's (1988) *Group Psychotherapy In Britain Today* (Open University Press) give a general overview of UK practice in the past. There are three good texts that introduce the group-analytic approach. The most recent by H, Behr and L. Hearst (2005) is *Group Analysis: A Meeting of Minds* (John Wiley). W. Barnes, S. Ernst and K. Hyde have written a recent overview, (1999) *An introduction to groupwork: a group-analytic perspective* (Palgrave Macmillan). And an earlier text by D. Kennard (1993) *The Workbook of Group Analysis* (Routledge) provides a clinically grounded study of practitioners at work. The range of other books in the International Library of Group Analysis and the journals, *Group Analysis, Group, and The International Journal of Group Psychotherapy* give access to the many specialist applications discussed here.

References

1. Burlingame, G.M., Mackenzie, K.R., Strauss, B. (2005). Small-group treatment. Evidence for effectiveness and mechanisms of change. Ch.14. In *Bergin and Garfield's Handbook of psychotherapy and behaviour change*. (ed. M.J.Lambert), New York. Wiley.
2. Foulkes, S.H. and Anthony, J. (1957*). Group psychotherapy: the psychoanalytic approach.* Penguin, Harmondsworth. (Maresfield Reprint, Karnac Books London 1989.
3. Yalom, I. and Leszcz, M. (2005). *The theory and practice of group psychotherapy.* Basic Books, New York.
4. Agazarian, Y. (1997). *Systems-centred therapy for groups*. Guilford Press, New York.
5. Creek, J. (ed.) (1997). *Occupational therapy and mental health*. Churchill Livingstone, Edinburgh.
6. Stone, W. (1996). *Group psychotherapy for people with chronic mental illness*. Guilford, New York.
7. Kanas, N. (1999). Group therapy with schizophrenic and bipolar patients. In G*roup psychotherapy of the psychoses* (eds. V. Schermer and M. Pines), pp. 129–47. Jessica Kingsley, London.
8. Schlapobersky, J. (1993). The language of the group: monologue, dialogue and discourse in group analysis. In *The psyche and the social world: developments in group- analytic theory* (eds. D. Brown and L. Zinkin), pp. 211–31. Routledge, London.
9. Schlapobersky, J. (1996). A group-analytic approach to forensic psychotherapy: from the speech of hands to the language of words. In *Forensic psychotherapy: crime, psychodynamics and the offender patient. Vol. 1: Mainly theory* (ed. C. Cordess and M. Cox), pp. 227–43. Jessica Kingsley, London.
10. Abse, W. (1979). Trigant Burrow and the inauguration of group analysis in the USA. *Group Analysis,* **3**, 218–29.
11. Lewin, K. (1951). *Field theory and the social sciences.* Harper and Rowe, New York.
12. Trist, E., Murray, H. (1990). T*he social engagement of social science Vol. 1, section 1.* University of Pennsylvania Press, Philadelphia, PA.
13. Pines, M. (1991). A history of psychodynamic psychiatry in Britain. In *Textbook of psychotherapy in clinical practice* (ed. J. Holmes), pp. 75–86. *C*hurchill Livingstone, Edinburgh,
14. Harrison, T. (2000). *Bion, Rickman, Foulkes and the Northfield Experiment: Advancing on a different front.* Jessica Kingsley, London.
15. Elliott, A. (1999). *Social theory and psychoanalysis in transition. Chapter 2, Free Association Books,* London. pp. 46–76.
16. Foulkes, S.H. (1948). *Introduction to group analytic psychotherapy.* Heinemann, London. (Maresfield Reprint, Karnac Books, London 1991.)
17. Dick, B. M. (1975). A ten year study of out-patients analytic group therapy. *British Journal of Psychiat*ry **127**, 365–75. Lorentzen, S. (2000) An assessment of change after long-term psychoanalytic group treatment. *Group Analysis* (in press).
18. Sigrell, B. (1992). The long-term effects of group psychotherapy. A thirteen year follow up study. *Group Analysis,* **25**, 333–52.
19. Lorentzen, S. (2000). An assessment of change after long-term psychoanalytic group treatment. *Group Analysis*, **33**.
20. Bion, W.R. (1961). E*xperiences in groups*. London: Tavistock.
21. Malan, D. (1976). A follow up study of group psychotherapy. *Archives of General Psychiatry* **33**, 1303–15.
22. Moreno, J. L. (1953). *Who shall survive? Foundations of sociometry, group psychotherapy and psychodrama*. Beacon House, New York.
23. Slavson, S. (1940). Group psychotherapy. *Mental Hygiene,* **24**, 36–49.
24. Wolf, A. (1949). The psychoanalysis of groups—1. *American Journal.of Psychotherapy,* **3**, 525–58.
25. Wolf, A. (1950). The psychoanalysis of groups—2. *American Journal of Psychotherapy,* **3**, 16–50.
26. Yalom, I. (1983). *In-patient group psychotherapy*. Basic Books, New York.
27. Rutan, J.S. and Stone, W. (2007). P*sychodynamic group therapy* (Fourth Edition). Guilford Press, New York.
28. Kaplan, H.I. and Sadock, B.J. (eds) (1993). C*omprehensive group psychotherapy*. Williams and Wilkins, Baltimore, MD.
29. Alonso, A. and Swiller, H.I. (eds) (1993). *Group therapy in clinical practice.* American Psychiatric Press, Washington, DC.
30. Ashbach, C. and Schermer, V. (1987). *Object relations, the self and the group.* Routledge, London.
31. Spotnitz, H. (1961). T*he couch and the circle*. Knopf, New York.
32. Ormont, L.R. (1992). *The group therapy experience: from theory to practice.* St. Martin's Press, New York.

33. Tubert-Olander, J. and de Tubert, R.H. (2004). *Operative Groups.* Routledge, London.
34. Janunsen, P. (1994). *Psychoanalytic therapy in the hospital setting.* Routledge, London.
35. Koenig, K. and Lindner, W.V. (1994). *Psycho-analytic group therapy.* Jason Aaronson, New York.
36. Tschuschke, V. (2000). The P.A.G.E. study: early treatment effected of long-term outpatient group therapies-first preliminary results. *Group Analysis*, **33**, 3, 397–411.
37. Tsegos, I.K. (1995). Further thoughts on group-analytic training. *Group Analysis,* **28**, 313–26.
38. S. Lorentzen (1990). Block training in Oslo: the experience of being both organiser and participant in the Norwegian Psychiatric Association group psychotherapy training programme. *Group Analysis*, **23**, 361–82.
39. May, R. and Yalom, I.D. (1989). Existential Psychotherapy. In *current psychotherapies* (eds. R. Corsini and D. Wedding.) pp. 363–402. F.E. Peacock, Itasca, Ill.
40. Bion, W. (2000). *Experiences in Groups*. Routledge, London (First published by Tavistock, 1961).
41. Ezriel, H. (1950). A psycho-analytic approach to group treatment. *British Journal of Medical Psychology*, **23**, 56–74.
42. Horwitz, L. (1977). A group-centred approach to group psychotherapy. *International Journal Group Psychothererapy*, **27**, 423–39.
43. Whitaker, D.S. and Lieberman, M.A. (1964). P*sychotherapy through the group process*. Tavistock, London.
44. Pines, M., Hearst, L. and Behr, H. (1982). Group analysis (group analytic psychotherapy). In *basic approaches to group psychotherapy and group counselling.* (ed. G. Gazda), pp. 132–178. Charles C. Thomas, Springfield Ill.
45. Pines, M. (1982). Reflections On Mirroring: 5th. Foulkes Lecture. *Group Analysis,* **15** (Supplement). Reprinted in Pines, M. (1998) *Circular reflections: selected papers on group analysis and psychoanalysis*. Jessica Kingsley, London, pp. 17–39.
46. Agazarian, Y., Peters, R. (1981). *The visible and invisible group: two perspectives on group psychotherapy*. Routledge, London.
47. Foulkes, S.H. (1964). *Therapeutic group analysis*. Allen and Unwin, London. (Maresfield Reprint, Karnac, London1984).
48. Pines, M. and Hearst, L. (1993). Group Analysis in *Comprehensive Group Psychotherapy.* (eds Kaplan I. and Sadock, J.) Williams & Wilkins, Baltimore.
49. Friedman, R. (2008). Dreamtelling as a request for containment: three uses of dreams in groups. *International Journal of Group Psychotherapy*, **58**, 3, 327–44.
50. Lipgar, R. and Pines, M. (2003). *Building On Bion: Vol. 1: Roots; Vol. 2*: Branches. Jessica Kingsley, London.
51. Nitsun, M. (2006). *The group as an object of desire: exploring sexuality in group therapy.* Routledge, London.
52. Cox, M. and Theilgaard, A. (1987). *Mutative metaphors in psychotherapy: the Aeolian mode.* Tavistock Press, London.
53. Bennis, W.G. and Shepard, H.A. (1956). A theory of group development. *Human Relations,* **9**, 415–37. Reprinted in *Sensitivity training and the laboratory approach (eds. R. Golembiewski and A. Blumberg)* pp. 91–115. *F.E. Peacock, Itasca, Ill.1970.*
54. Foulkes, S.H. (1964). *Therapeutic group analysis.* George Allen and Unwin, London. pp. 51–2, 176–77.
55. Foulkes, S.H. (1964). *Therapeutic group analysis*. George Allen and Unwin, London. p. 57.
56. Tschuschke, V. and Dies, R.R. (1994). Intensive analysis of therapeutic factors and outcome in long-term inpatient groups. *International Journal of Group Psychotherapy,* **44**, 185–208.
57. Marziali, C. et al. (1997). The contributions of group cohesion and group alliance to the outcome of group psychotherapy. *International Journal Group Psychotherapy*, **47**, **4**, 475–97.
58. Pines, M. (1986). Coherence and its disruption in the development of self. *British Journal of Psychotherapy*, **2**, **3**, 180–85. Reprinted in Pines, M. (1998) *Circular Reflections: selected papers on group analysis and psychoanalysis,* Chapter 12. Jessica Kingsley, London. pp. 211–23.
59. Grotjahn, M. (1977). *The art and technique of analytic group therapy.* Jason Aaronson, New York. p. 14.
60. Pines, M. (1990). Group analysis and the corrective emotional experience: is it relevant? *Psychoanalytic Inquiry*, **10**, **3**, 389–408.
61. Garland, C. (1982). Group analysis: taking the non-problem seriously. *Group Analysis,* **15**, 4–14.
62. Pines, M. (1995). The universality of shame: a psychoanalytic approach. *British Journal of Psychotherapy*, **11**, 346–57.
63. Nitsun, M. (1996). *The anti-group: destructive forces in the group and their creative potential.* Routledge, London.
64. Salvendy, J.T. (1993). Selection and preparation of patients and organisation of the group. In *Comprehensive group psychotherapy* (ed. H.I. Kaplan and B.J. Sadock). Williams and Wilkins, Baltimore, MD.
65. Hearst, L. (1998). The restoration of the impaired self in psychoanalytic treatment. In *Borderline and narcissistic patients in treatment* (ed. N. Slovinska-Holy) International University Press, New York. Chapter 7.
66. Roback, H.B. and Smith, M (1987). Patient attrition in dynamically oriented treatments in groups. *American Journal of Psychotherapy*, **144**, 426–31.
67. Melnick, J. and Woods, M. (1976). Analysis of group composition: research and theory for psychotherapeutic and growth-oriented groups. *Journal of Applied Behavioural Science*, **12**, 493–512.
68. Thygesen, B. (1992). *Diversity as a group-specific factor. Group Analysis* **25**, 175–86.
69. Zelakowski, P. (1998). The sub-optimal group. *Group Analysis*, **31**, 491–504.
70. Stern, M.J. (1993). Group therapy with medically ill patients. In *Group therapy in clinical practice* (ed. A. Alonso and H.I. Swiller), pp. 185–200. American Psychiatric Press, Washington, DC.
71. Leszcz, M. and Goodwin, P. (1998). The rationale and foundation of group psychotherapy with metastatic breast cancer. *International Journal of Group Psychotherapy,* **48**, 245–73.
72. Spiegel, D., Bloom, J.R. and Yalom, I.D. (1981). Group support for patients with metastatic cancer. *Archives of General Psychiatry*, **38**, 527–33.
73. Spiegel, D. (1999). *Supportive group therapy with cancer patients.* Basic Books, New York.
74. Kanas, N. (1999). Group therapy with schizophrenic and bipolar patients. In *Group psychotherapy of the psychoses* (eds. V. Schermer and M. Pines) Jessica Kingsley, London.
75. Kapur, R. (1993). The effects of group interpretations with the severely mentally ill. *Group Analysis,* **26**, 411–32.
76. Resnik, S. (1987). *The theatre of the dream.* The New Library of Psychoanalysis, London; and (1995) *Mental Space*. Karnac. London.
77. Chiesa, M. and Fonagy, P. (2000). The Cassel Hospital personality disorder study. *British Journal of Psychiatry (forthcoming).*
78. Wilborg,T. and Karterud, S. (2001). The place of group psychotherapy in the treatment of personality disorders. *In Current Opinion in Psychiatry,* **14/2**, 125–30.
79. Karterud, K., Oyvind, U. (2004). Short-term day programmes for patients with personality disorders. What is the optimal composition? *Nordic Journal Psychiatry,* **58**(3) 243–49.
80. Cox, M. (1986). The 'holding function' of dynamic psychotherapy in a custodial setting: a review. *Journal of the Royal Society of Medicine*, **79**, 162–64.
81. Kennard, D. (1993). Group therapy at Rampton. Group Work Co-ordinating Committee, Rampton Hospital.
82. Genders, E. and Player, E. (1995). *Grendon: A study of a therapeutic prison*. Clarendon, Oxford University Press.

83. Welldon, E. (1994). Forensic Psychotherapy. In *Handbook of psychotherapy*, (eds. P. Clarkson and M. Pokorny), pp. 470–93. Routledge, London.
84. Lees, J., Manning, N. and Rawlings, B. (2000). Therapeutic community effectiveness: A systematic international review of therapeutic community treatment for people with personality disorders and mentally disordered offenders. *CRD Report 17: University of York NHS Centre for Reviews and Disemmination.*
85. Brown, A., Caddick, B. (eds) (1993). *Groupwork with offenders*. Whiting and Birch, London. p. 2.
86. Welldon, E. and Wilson, P. (2006). Special edition of Group Analysis, Forensic Psychotherapy. *Group Analysis*, 39.1.
87. Van der Kolk, B. (1993). Groups for patients with histories of catastrophic trauma. In *Group therapy in clinical practice* (ed. A. Alonso and H.I. Swiller). American Psychiatric Press, Washington, DC. pp. 289–305.
88. Hall, Z., Mullee, M. and Thompson, R. (1995). A clinical and service evaluation of group therapy for women survivors of childhood sexual abuse. In *Research Foundations for Psychotherapy Practice* (eds. M. Aveline and D.A. Shapiro). Wiley, Chichester.
89. Lifton, R. (1983). *The broken connection*. Basic Books, New York.
90. Greene, L.R. (ed.) (2005). Special Edition of the International Journal of Group Psychotherapy: Children & Adolescents in the Aftermath of 9/11: Group Approaches Towards Healing, Trauma and Building Resilience, July 2005.
91. Weinberg, H. and Nuttman-Shwartz, O. (eds) (2005). Special Edition of Group Analysis on Trauma, Vol. 38 No 2, June 2005.
92. Schlapobersky, J. and Bamber, H. (1988). Rehabilitation with victims of torture. In *Refugees—The trauma of exile*. pp. 206–22. Nijhoff, The Hague.
93. Woodcock, J. (1997). Groupwork with refugees and asylum seekers. In *Race and groupwork*, (eds. T. Mistry and A. Brown), pp. 254–77. Whiting and Birch, London.
94. Callaghan, K. (1996). Torture—the body in conflict: The role of movement psychotherapy. In *Arts Approaches to Conflict*, (ed. M. Liebmann). Jessica Kingsley, London.
95. Benun, I. (1986). Group Marital Therapy: A Review. *Sexual and Marital Therapy*, **1**, 61–74.
96. Skynner, A.C.R. (1986). Recent developments in marital therapy. In *Explorations with families: group analysis and family therapy: selected clinical papers of Robin Skynner*. Routledge, London.
97. Lesecz, M. (1997). Integrated group psychotherapy for the treatment of depression in the elderly. *Group* **21**, 89–113.
98. Dies, R.R. (1993). Research on group psychotherapy: overview and clinical applications. In *Group therapy in clinical practice*, (eds. A. Alonso and H.I. Swiller), pp. 475–6. American Psychiatric Press, Washington DC.
99. Roback, H. (2000). Adverse outcomes in group therapy. *Journal of Psychotherapy Practice and Research*, **9**, 113–22.
100. Robinson, L.A., Berman, J.S. and Niemeyer, R. (1990). Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research. *Psychological Bulletin,* **108**, 30.
101. Tyllitski, C.J. (1990). A meta-analysis of estimated effect sizes for group versus control treatments. *International Journal of Group Psychotherapy,* **40**, 215–24.
102. Piper., W.E., Perrault, E.L. (1989). Pre-therapy preparation for group members. *International Journal of Group Psychotherapy* **39**, 17–34.
103. Lorentzen, S. and Heglund, P. (2004). Predicting change in long-term group psychotherapy. *Psychotherapy and Psychosomatic Medicine,* **73**, 1. 125–135; and Lorentzen, S. (2006) Contemporary Challenges For Research. In *Group Analysis*, 3, 321–340; and Mace, C., (2006) Setting the world on wheels: some clinical challenges of evidence-based practice. In *Group Analysis*, 39, 3, 304–20.
104. Dies, R. R. (1993). Research on group psychotherapy: overview and clinical applications. In *Group therapy in clinical practice*, (eds. A. Alonso and H.I. Swiller), p. 502. American Psychiatric Press, Washington DC.
105. Tschuschke, V. and Dies, R.R. (1994). Intensive analysis of therapeutic factors and outcome in long-term inpatient groups. *International Journal of Group Psychotherapy*, **44**, 183–214.
106. Tschuschke, V., Anbeh, T., Kiencke, P. (2007). Evaluation of long-term analytic outpatient group therapies. In *Group Analysis,* **40**, **1**, 140–59.
107. Lorentzen, S., Bogwald, K. and Hogland, P. (2002). Change during and after long-term analytic psychotherapy. *International Journal of Group Psychotherapy*, **52**(3) 419–29.
108. Terlidou, C., Moschonas, D., Kakitsis, P., Manthouli, M., Moschona, T., Tsegos, I. (2004). Personality changes after the completion of long-term group psychotherapy. *Group Analysis,* **37**, **3**, 401–18.
109. Conway, S., Audin, K., Barkham, M., Mellor-Clark, J. and Russell, S. (2003). Practice based evidence for a brief time-intensive multi-modal therapy guided by group-analytic principles and methods. *Group Analysis* **36**, **3**, 413–35.
110. Schamess, G. (1994). Group psychotherapy with children. In *Comprehensive group psycho*therapy (ed. H.I. Kaplan and B.J. Sadock). Williams and Wilkins, Baltimore, MD. pp. 560–77.
111. Kymissis, P. Group psychotherapy with adolescents. In *Comprehensive group psychotherapy* (ed. H.I. Kaplan and B.J. Sadock), pp. 577–84. Williams and Wilkins, Baltimore, MD.
112. Evans, J. (1998). *Active analytic group therapy for adolescents*. Jessica Kingsley, London.
113. Melzak, S. (ed.) (2000). *Children in exile: therapeutic and psychotherapeutic work in the clinic and the community.* Jessica Kingsley, London.
114. Heinzel, R. (2000). Outpatient psychoanalytic individual and group psychotherapy in a nationwide follow-up study in Germany. *Group Analysis*, **33**, 353–72.

6.3.7 Psychotherapy with couples

Michael Crowe

Introduction and background

There is an ongoing crisis in the institution of marriage, at least in Western cultures. There has for some time been a tendency to idealize marriage, and at the same time social forces are operating which tend to undermine it.(1) These influences have probably made a contribution to the increasing divorce rate, as well as the tendency for fewer couples to marry, and have probably also led to an increase in the number of couples seeking help with their relationships.

In the United Kingdom, for example, the number of marriages taking place each year has fallen for the first time in living memory, and the number of divorces is still steadily increasing, reaching 40 per cent of marriages in 1996.(2) There are also a large number of 'common-law' marriages, often with children, as well as more transient cohabiting or non-cohabiting sexual relationships, both heterosexual and homosexual. The stability of these relationships is, of course, not recorded in the marriage or divorce statistics, and the rate of breakup can only be guessed at; however, it is very probable from clinical experience that in these non-marital relationships there is a higher than 40 per cent incidence of breakup. In the wake of these changes, there are a large number of

single-parent families and 'reconstituted' or blended families, as reviewed by Robinson,[3] and there is a decreasing proportion of children who are being brought up in the traditional nuclear family with two biological parents.

In addition to these new factors affecting marriage in the early 21st century we should also be aware of the fact that many countries, especially those in the developed world, have a multicultural society, and that immigrant cultures have different attitudes to marriage and family life. For example, families from the Indian subcontinent often prefer to arrange marriages for their children, and in some cases insist that the couple live in the husband's parents' house. On the other hand, West African couples often leave their children in Africa to be looked after by family members for long periods of time, while the parents work or study in the West.

In the last few years, Gay marriage or Civil Partnership has been recognized in many western countries. Couples in Gay relationships have many of the same problems and satisfactions as heterosexual couples, and in addition, must live with fairly widespread negative attitudes and homophobia from neighbours, family and society generally. Their relationships have to be, if anything, stronger than heterosexual ones to survive these pressures, and may be more in need of therapy.

Couple therapy must be able to take account of these factors, and whilst much of what is contained in this chapter will relate to heterosexual married British couples living with their biological children, it should be understood that there are many other types of relationship which can be helped using a similar approach, with appropriate changes of emphasis. In a later section, there will be some additional discussion of the specific problems relating to couples from other cultures, and ways of managing these.

Couple counselling and couple therapy

The concept of couple counselling dates from the 1920s when in the United States the American Association for Marital Counselling was formed; in the United Kingdom the Marriage Guidance Council (now called Relate) was founded in 1938. Counselling mainly took the form of giving advice on practical issues, but in more recent years, Relate counselling has been orientated more towards psychodynamic approaches, and favours a longer-term involvement with the couple. Couple counselling continues in both countries, and the great majority of couples seeking help with their relationships are seen by couple counsellors, rather than any other types of therapist.

The distinction between couple counselling and couple therapy is not an easy one, because many of the interventions are similar. In a simplistic sense therapy attempts to make a more radical difference to the couple's functioning than counselling, which has the general aim of improving adjustment to the situation as it is. However, many forms of both couple therapy and couple counselling are based on a theoretical formulation which is derived from a related school of individual psychotherapy (for example, cognitive behavioural or psychodynamic). Thus, theoretical formulations in the resultant couple work are so different between therapies (e.g. psychodynamic as against behavioural) that a particular form of counselling may have more in common with a related form of couple therapy than that therapy itself has with another type of couple therapy.

Psychoanalytic/psychodynamic couple therapy

Couple therapy using a psychodynamic model began in the United Kingdom in 1948, when Dicks and his colleagues founded the Institute of Marital Studies. The theories and techniques involved have been ably reviewed by Daniell[4] and Clulow.[5] The central concept used is that the inner (unconscious) world of the two partners determines their interaction and their response to changing circumstances. It is as though each partner has an internal blueprint, both of themselves and each other, formed partly by observation but also partly by the influence of earlier intimate attachment experiences with parents, siblings, or friends. These influences may actually determine the choice of partner, and the nature of each partner's patterns of attachment (secure or insecure) will affect the ways in which they cope with the stresses of the new relationship. There may then be projections which lead one partner to attribute motives such as hostility or sadism to the other, whereas in fact this is a split-off and denied characteristic of the first partner. Other consequences of this unconscious process may include the system of shared fantasies and defences which builds up as the relationship continues.

In therapy, four premises are used, which inform a relatively long-term and open-ended series of sessions. The first is that a person's emotional health is related to his or her capacity to manage both internal conflict and external stress: it is important to be able to experience fear as well as trust, pain as well as pleasure, doubt as well as certainty, frustration as well as satisfaction. Secondly, significant relationships can be used to resurrect, but also change, inflexible patterns of behaviour established in the past. Thirdly, unconscious processes need to be taken into account when attempting to understand problems in relationships. Fourthly, change takes time because it requires a reordering of perceptions of self and others, perhaps with the help of transference interpretations by the therapist involving both partners.

Therapy in this mode may be carried out by one therapist seeing both partners, but is more often done by two therapists either seeing the couple together or in parallel individual sessions, using one partner with one therapist, with joint supervision of the two therapists. An intriguing aspect of this therapeutic format is that sometimes the two cotherapists find themselves interacting in unfamiliar ways, in sessions and between sessions, which are thought to represent the projection of fantasies and feelings by the couple on to the therapists; the therapists' understanding of these projections in their joint supervision may play a role in advancing the therapy itself. If these insights are used to inform the therapists' interaction with the couple, the individual partners may then be made aware of their own conflicts, fantasies, and projections, and thus be able to give up some of their repetitive patterns of behaviour and withdraw damaging projections.

The psychoanalytic approach has been an important source of theoretical ideas in couple therapy, especially the concepts of attachment and loss developed by Bowlby.[6] It has also the distinction of being the first theory to be adapted to this area of work. There are, however, some drawbacks to working in this way, as enumerated by Wile.[7] He sees the emphasis on negative impulses and emotions (e.g. dependence, narcissism, sadism, manipulation, and exploitation) as painting a rather unflattering and negative picture of the couple in therapy, and perhaps therefore reducing

their motivation to continue. A more serious problem with the approach is that the psychodynamic concepts, whether of defence mechanisms, projections, or shared fantasies, are treated as if they were as real as observed behaviour, whereas in fact they must remain assumptions based on hypothetical constructs, and are really only valuable in so far as the therapy based on them is effective.[1]

The question of efficacy is raised later in the chapter, but it must be stated here that the psychodynamic therapies for couple problems have only seldom been submitted to controlled trial, and then usually in a relatively short-term form. The therapy may be quite long term, and the improvements seen are usually not dramatic, so that in the last analysis the approach has to remain of uncertain value.

Behavioural couple therapy

The behavioural approach, in contrast, makes no assumptions about internal conflicts or underlying mechanisms in the individuals. The approach was initiated in 1969 by Stuart[8] and Liberman[9] as behavioural marital therapy. They worked from the principles of operant conditioning and made the assumption that couples who were having difficulties were either giving each other very low levels of positive reinforcement or were using punishment or negative reinforcement to coerce each other into behaving differently. The remedy that they proposed for this situation was to help the partners to learn how to persuade each other to conform to the desired pattern of behaviour by the use of prompting and positive reinforcement. Thus, complaints would be transformed into requests and requests into tasks agreed by both partners.

Behavioural marital therapy relies on the therapist's observation of the couple's behaviour in the session and on the problems they report from the previous week or equivalent timespan. There are two types of therapeutic activity in behavioural marital therapy. The first is reciprocity negotiation, in which the partners request changes in behaviour on each side and negotiate how this can be achieved through mutually agreed tasks. The second is communication training, in which the partners are encouraged to speak directly and unambiguously to each other about feelings, plans, or perceptions, and to feed back what they have heard and understood. In both these approaches, the deeper meanings behind a particular piece of behaviour are ignored, the emphasis being on change in the interaction both in the here and now and in the immediate future. The approach has been the subject of many controlled trials (see below), and is of proven efficacy.

Cognitive behavioural and rational–emotive couple therapy

Aaron Beck,[10] in his cognitive behavioural approach to couple therapy, identifies in the communication of disturbed couples many of the problems found in the thinking of depressed patients, and attempts to correct these. Thus, he tackles misunderstandings, generalizations, untested assumptions, and automatic negative thoughts by challenging assumptions, reducing unrealistic expectations, relaxing absolute rules, improving the clarity of the communication and focusing on the positive rather than the negative.

Similarly, Albert Ellis (reviewed by Dryden[11]) uses a rational-emotive approach to couple problems. Here, the main focus is on the use of words; terms such as 'intolerable' are replaced by (for example) 'difficult to accept', and the couple are encouraged to express desires rather than demands. There is an analysis of the repetitive cycles of cognitive and behavioural disturbance, in which each partner may attribute the other's behaviour to a negative motive and assume that nothing can be done about it. The general thrust of this therapy is similar to that of Beck, but with a more lively and less formalized approach in the session.

Systems therapy for couple problems

The systems approach to couple therapy derives partly from concepts developed by Minuchin[12] and Haley,[13] and partly from the work of Selvini Palazzoli *et al.*[14] All these pioneers worked predominantly with families rather than couples, but many of their ideas and techniques are relevant to the treatment of couples. Although the systems approach to therapy has broadened and deepened since the 1980s, many of the early concepts are still very useful.

A central concept in thinking about couple relationships is 'enmeshment', by which is meant an excessive involvement in what is essentially the private business of another person. It is quite common to find an enmeshed relationship between parents and their teenage children, in which both sides find it very hard to 'let go'. It can also be found in couple relationships where one partner wants to be closer than the other, and a conflict arises as to what is the best distance to maintain. Systems therapy aims to help them to find a compromise 'distance' which suits them both, and thereby to reinforce the necessary 'boundaries' which people need in maintaining their individuality within a relationship.

The concept of circular causality is also central to systems work. This enables the couple to get away from the idea that one person is necessarily to blame for a particular situation by considering the continuous cycle of cause and effect in which A's actions may be caused by B's actions and also B's may equally be caused by A's. Thus, systems therapists, when approaching a couple problem, do not focus on one partner's behaviour, but rather on the pattern of interaction obtaining in the relationship. They will then try to effect a change in which both partners contribute actively to the solution of the problem.

Systems therapists have many techniques at their disposal, including those which increase the couple's understanding of the system they are participating in. These include family genograms (a form of family tree construction which leads to discussion of transgenerational influences or 'systems over time'), family 'sculpting' (in which the members position themselves and each other wordlessly to represent their current relationships), and the discussion of 'family myths' and stories. More active techniques, designed to play a part in changing the family interaction, include creating conflict in the session, giving homework tasks, and the use of 'paradoxical injunctions' in which the therapist tells the family to continue with the current interaction because, even though it is problematic, it seems to be protecting them from worse consequences. These more active techniques will be dealt with in more detail in the main part of this chapter on behavioural–systems therapy.

Mixed or eclectic approaches

Most couple therapists use a mixture of techniques, and it seems that this is probably an inevitable consequence of the difficulties

involved in applying one therapeutic method rigorously in a clinical setting. A number of specific combinations have been advocated, and will be briefly mentioned here.

The first is the psychodynamic-behavioural approach of Segraves.[15] In this, the basic underlying cause of marital disturbance is assumed to be the partners' conflicting internal and unconscious projections, and their interactions. The therapy, however, is not only directed at helping them to understand these (as in psychodynamic therapy) but also to increase their negotiating and communicating skills (as in behavioural marital therapy).

The second is a more comprehensive mixture of theory and technique, known as the intersystem model, and advocated by Weeks.[16] This tries to take account of the individual, interactional, and intergenerational aspects of couple relationships, and combines them in what is probably closest to a systems model, but with more emphasis on the psyche of the individual. Interventions are on both a conjoint and individual basis, and the techniques of decentring (see below) and paradoxical injunctions are often used.

The third eclectic approach is that of Spinks and Birchler.[17] This is called behavioural–systems marital therapy, and makes use of behavioural marital therapy as the main form of intervention, moving into the systems mode when 'resistance' emerges. There are many similarities between this form of treatment and the one described in the main part of this chapter, but our 'behavioural–systems approach' is more integrated as between the two components of the method.

The fourth eclectic approach which should be mentioned is that of Berg-Cross.[18] She uses rational-emotive, sociocognitive, systemic, psychodynamic, humanistic, and theological concepts to understand and modify couple relationships. Like that of Weeks, her approach gives the therapist a very wide canvas to work on, but may lose some of the focus by being very general and all-embracing.

The behavioural–systems approach to couple therapy

Behavioural–systems couple therapy is the approach that will be described in detail in the present chapter, and although it is only one of several approaches to couple problems, it has the advantage of spanning two of them, and issues such as indications for therapy and assessment are shared with both the pure behavioural and systemic approaches. It is the method developed at the Maudsley Hospital Couple Therapy Clinic in the 1980s. It has been expounded at greater length by Crowe and Ridley,[1] and like some of the other eclectic models mentioned above it combines two different approaches, behavioural marital therapy and systems family therapy. The behavioural dimension, similarly to that described by Stuart (8) and Jacobson and Margolin,[19] consists of the relatively straightforward methods of reciprocity negotiation and communication training. The systems dimension is more complicated, and involves systems thinking, structural moves during the session, tasks and timetables for the couple between sessions, and the use of paradox. The method was developed in a predominantly psychiatric setting, and has been found to be particularly suitable for those couples where one or both partners has psychiatric problems in addition to their relationship difficulties. It is also useful as an adjunct to psychosexual therapy where a sexual dysfunction or a sexual motivation problem seems to be connected with relationship issues.

The method should be thought of as a series of menus from which the practitioner can choose techniques rather than as a set course of therapy beginning at one point and ending at another. Thus, the various components of behavioural–systems couple therapy can be incorporated at any time in the therapy session, although in practice, negotiation, communication training, and structural moves are usually employed in the earlier part of the session, while tasks, timetables, and paradox are usually reserved for the 'message' at the end, and are linked to homework assignments to be carried out between sessions.

The different techniques of behavioural–systems couple therapy can be thought of as belonging to a kind of hierarchy. The so-called 'hierarchy of alternative levels of intervention (ALI)'[1] links each type of intervention with a particular set of clinical problems and makes recommendations as to the type of intervention that is appropriate. The ALI hierarchy is shown in diagrammatic form in Fig. 6.3.7.1. As may be seen, where the couple appear to have greater rigidity in their behaviour, where they show more symptoms, and where they show more reluctance to accept the relationship as the focus of work, the therapist needs to move to the systems end of the hierarchy, and use more ingenuity in the development of interventions. If, however, the couple accept the interactional focus and show willingness to recognize the part that the relationship is playing in maintaining the problems, the therapist may be quite comfortable and effective working behaviourally. By and large, the preference is to work behaviourally, since this implies collaborating with the couple and accepting their stated goals, whereas the systems approach puts the therapist into a more managing role, deciding what is best for the couple and suggesting tasks that may not be what they would expect. It should be emphasized that the therapist may at any stage move up or down the hierarchy, according to the couple's response: an increase in flexibility shown by the couple could be the trigger for the therapist to begin working in a more behavioural way, whereas an increase in rigidity or a failure to respond to behavioural work could be met by a more systemic approach.

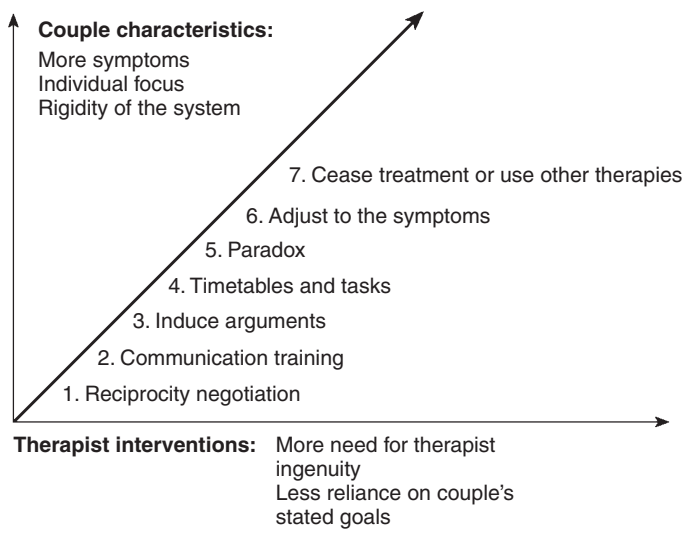


Fig. 6.3.7.1 The alternative levels of intervention hierarchy.

Indications and contraindications

If there is a relationship problem identified by either the couple or their advisers, even if there are also individual psychiatric or behavioural problems, and if the couple are willing to attend together, then in most cases they are suitable for behavioural–systems couple therapy. The breadth of the therapeutic approach, the fact that the behavioural techniques are of proven efficacy (see below), and the fact that the systemic interventions are suitable for those with more psychiatric symptoms or similar problem behaviours, all give the therapy a wide range of positive indications.

(a) The nature of the problem

Clearly those with relationship problems such as arguments and tensions are highly suitable for couple therapy. Another related indication is those relationships in which one partner (who might be attending a counsellor or psychiatrist alone) spends much time complaining about the absent partner's behaviour. A third indication is where the health of one partner suffers following the other partner's individual therapy.

Many problems with sexual function would be suitable for couple therapy, including those couples where there is a disparity in sexual desire, or those where one partner has a specific phobia for sex. In some such cases there is also a need for individual therapy, especially where one partner is the survivor of earlier childhood sexual abuse.

Many people with depression or anxiety, especially those where there is also poor self-esteem, may be suitable for couple therapy. There are often aspects of the illness that are exacerbated by problems in the relationship. Indeed, in a study by Leff *et al.* in 2000[20] it was shown that couple therapy was an effective and acceptable form of treatment for couples in whom one partner was depressed.

Where jealousy is present the problem usually affects the non-jealous partner to a greater or lesser extent, and here it would almost always be useful to have at least a few conjoint sessions with the couple, as suggested by De Silva.[21]

Some problems are perhaps less amenable to couple work, and among these are, for example, phobias which seem unconnected with home life in any way, and post-traumatic stress reactions where the event happened away from the partner. Some alcoholic and drug-addicted patients have so much of their existence involved with the addiction that they are not available emotionally to do couple work, and the work would at that stage be wasted on them. Similarly, those with an acute psychosis would, at the time they are acutely ill, be unavailable to this kind of therapy, and should not be offered it. However, in both cases, when the acute crisis is over and the addiction or psychosis is under control, it would be very appropriate to offer them some kind of couple therapy, even if this had limited aims and expectations. Some of the most useful psychological interventions in schizophrenia, after the acute illness has resolved, involve the nearest relative, as shown by McFarlane.[22]

(b) Degree of connection with the relationship

Some problems in individuals have been in existence long before they entered the present relationship. If this is the case, the therapist should consider whether it is best to embark on couple therapy or whether individual therapy would be better. However, even when there seems no causal connection with the relationship, the effect of the problem on the partner may be such as to warrant at least one or two couple sessions.[23]

(c) Availability and willingness for joint therapy

If the partner of a patient is unavailable or unwilling to attend for therapy, it may be appropriate to let the situation be, and not offer treatment. In some immigrant couples, for example those from the Indian subcontinent, there are cultural reasons given for a wife not attending therapy, and we usually have to respect these. However, in both situations it is sometimes right to put some pressure on the absent partner to attend, because the reasons for non-attendance may be relevant to the couple problems we are trying to treat.

(d) Is the relationship continuing?

If the person who is asking for therapy is going through a divorce or equivalent breakup of a relationship, it may not be appropriate or possible to treat both partners. However, in some cases there is good work to be done in arranging a more satisfactory breakup, in terms of domicile and care of any children. This 'mediation' work is increasingly being done, and many of the processes are similar to those of couple therapy.

Assessment and selection

This is not an easy process, because there is a dearth of research on the types of couple problem presenting for treatment and on the outcome of treatment itself. However, on the basis of the referral letter (usually from a GP or psychiatrist) the therapist will usually arrange a preliminary session with the couple. The only reason not to see the couple in the first instance is that they are not willing to attend as a couple, or the problem is seen clearly as an individual one (see above).

Another, more subtle, form of selection occurs during the first therapy session, when the couple are in contact with the process of therapy. In this session, the therapist considers the ability of the partners to empathize with each other, their pattern of communication, their stage in the 'family life cycle' (see below) with its associated stresses, and their flexibility in response to simple therapeutic interventions. He or she will attempt to use reciprocity negotiation or other straightforward techniques of therapy, partly as a treatment trial, and partly to see whether the couple is ready for this kind of intervention. If not, the therapist can move to a more systemic approach, or decide after a few visits that there is no future at this point in further therapy of this kind, and that something else (e.g. individual therapy) is needed (Fig. 6.3.7.1).

The process of therapy: beginning and continuing

Behavioural–systems couple therapy is essentially a short-term therapy involving perhaps 5 to 10 sessions of 60 minutes each, over a period of 3 to 6 months. Although the therapy was developed in a room with a one-way screen and live supervision, it is quite possible to use behavioural–systems couple therapy in any conventional consulting room without a team or live supervision. If the one-way screen is used, the therapist and the team (which may include students as well as experienced colleagues) begin by reading the referral letter and the biographical questionnaires which the couple will have completed, and discuss the case with a view to formulating the problem from an interactional point of view. This may involve, for example, thinking about the couple's stage in the 'family life cycle' (e.g. birth of the first child or the 'empty nest'),

any recent event such as a bereavement or a new relationship, or the diagnosis of a serious illness. Hypotheses about the possible causation of the recent problems are not necessarily thought of as 'true' explanations, but have the function of informing the therapist's thinking and suggesting what level of the ALI hierarchy to choose at the initial meeting and what strategy to employ in the session.

There are several important issues to address in the first session, which functions both as a kind of assessment and as the beginning of the therapeutic work. It is necessary:

- to remain in control;
- to develop rapport with both partners, without favouring either;
- to maintain the momentum of the session and the interactional focus;
- to maximize the opportunities for the couple to experience a change in the nature of their interaction.

A particularly useful move at some stage in the first session, and one which we almost always use, is to ask the partners to talk directly to each other rather than the therapist; this is the so-called 'decentring' technique originated by Minuchin.[12] Keeping this configuration for as much as possible of the session enables the therapist to observe the couple's typical pattern of interaction, to intervene as a 'theatrical producer' rather than a diplomatic negotiator, to avoid as far as possible taking sides, and to encourage the kind of negotiation which hopefully the couple will be able to carry on at home without the presence of a third party. It is not necessary to remain decentred for the whole session, but if it does not happen at all in the first session, an opportunity for effective work will have been lost.

It may be difficult to remain decentred in the face of pressures from one or both partners to talk to the therapist directly. One way to participate while still remaining decentred is to request the partners to ask each other questions to which the therapist would like to know the answer. For example, one might say: 'Could you ask your partner what she thinks about your coolness on this matter?' or 'Perhaps your partner disagrees with you; could you check with him?'. In this way the couple can continue to talk to each other in the decentred position without undue 'triangling in' of the therapist.

One situation which causes particular problems for the behavioural–systems therapist is where one partner persists with monologues, either about his or her own symptoms and problems or about the partner's unreasonable behaviour. This is a perfectly acceptable way to present a problem in one-to-one therapy, but in couple therapy it slows down the interaction and prevents the therapist focusing on the relationship. One way to overcome this is by decentring, but in some couples this is too difficult, and another possibility is to ask the non-verbal partner to comment on the spokesperson's problems. This can provoke a minor crisis in the couple, and lead to the spokesperson realizing that the problems are not all one-sided. It is a technique related to circular questioning, and is used fairly extensively in family therapy see Chapter 6.3.8

Another obstacle to progress is the situation in which the partners have intractable arguments, perhaps about other family members. If such battles continue to dominate the sessions the therapist may have to devise a method for putting them 'on ice for the time being' and concentrating instead on negotiating everyday problems to do with the house or the children.

Momentum may also be slowed by the therapist's own style of working. Students of the behavioural systems approach may need to 'unlearn' some of their otherwise good therapeutic habits such as being a good and empathetic listener. They may be able to achieve this by decentring or by asking circular questions in order to refocus on the interaction and increase momentum.

On the other hand, it is still important for the therapist to be able to feel and show empathy to the two individuals in therapy. The difficulty may be that to show empathy to one partner may be interpreted by the other as side-taking. A possible remedy lies in a particular skill, which, however, is not easy to acquire, of saying something to show that one has understood one partner without antagonizing the other. One way to develop this skill is for the therapist not to become emotionally involved in the issues, but to concentrate on the process of interaction, thinking all the time in terms of balance and communication rather than worrying about the rights and wrongs of what is being discussed. This must be done while still showing respect to both the partners, taking their problems seriously, and at the same time conveying hope that they can be solved.

It is also very helpful for the partners themselves to be in touch with each other emotionally and take each other seriously. Some individuals are very good at communicating, but only at an intellectual level, and cannot express empathy with each other. Their interaction is like that of fellow committee members, and they tend to suppress any expression of feelings such as sadness or anger which are 'not on the agenda'. In other couples, there is an imbalance, with one partner expressing feelings openly and the other being exclusively logical and self-controlled. In both situations it is necessary to help them to communicate both intellectually and emotionally, encouraging the self-controlled individual to be more open and the person who is emotionally open to try to be more restrained at times.

Ending the session

Every session must have an ending, and the purpose in the behavioural–systems approach is to send the couple away with something to work at in the weeks before the next session. About two-thirds of the way through a supervized session the therapist will usually go behind the one-way screen, turning off the camera and closing the shutters (although when working alone this luxury is not available and the therapist must work independently to end the session on a positive note). The team and the therapist then spend about 15 minutes in discussion, planning the 'message' to be given to the couple at the end. Part of this discussion will centre round the team's thinking about the significance of the problems from a systems point of view, but part will also be concerned with how the therapist can help the couple to change their interaction.

The introduction to the message will give the date of the next appointment, and usually also contains some positive and sympathetic comments for both partners. It is important as far as possible to keep them both 'on side' at this stage, so that it would be unwise to say something which could be seen by either partner as favouring the other one. A good example of an introductory comment would be: 'The team and I are aware of the great difficulties you are experiencing, but we think you have what is basically a good relationship, and you are both working hard to improve things'.

The message itself will vary according to the level of the hierarchy at which the therapy is being pitched. If it is mainly a session of reciprocity negotiation the main theme may be simply to reiterate the negotiated plans for both partners which have emerged from the earlier discussion. If, however, the therapist is working more systemically, the message may contain a task, a timetable, or a paradoxical injunction which is designed to alter both the behaviour of the couple and the way they conceptualize their relationship. In some cases, it is appropriate to use a 'split-team' message , in which one part of the team is said to favour a more behavioural task while the other part believes that that will be impossible to achieve and therefore prefers to 'prescribe the symptom'.

The final part of the message is again likely to reiterate the positive sentiments of the introduction. There are good therapeutic reasons for this, in that people tend to remember the positive things that they hear about themselves, and may then link these to the more specific tasks or injunctions that are given with them. In many cases we also send a written copy of the message to the couple, so that they can think it over between sessions and not forget what has been discussed.

Where therapists are working alone, they will be unable to have the 15 minute break in the session, but it is always useful to spend a little time thinking about the message to be given at the end of the session, and to think what other team members might have said or suggested for a final message.

Specific techniques in behavioural–systems couple therapy

(a) Reciprocity negotiation

Within the alternative levels of intervention hierarchy, reciprocity negotiation is at the lowest level, relating closely to the goals that the couple themselves has set, and depending on a fairly co-operative attitude on both sides. The partners state their complaints in everyday terms, and the task of the therapist is then to help them to achieve a compromise by each doing what the other partner wants in a reciprocal way.

Reciprocity negotiation is partly based on operant conditioning and partly on the social exchange theory of Thibault and Kelley.[24] The assumption is that satisfaction in marriage and other intimate relationships is based on a relatively equal and high level of input by each partner of positive (i.e. rewarding) behaviour and a relatively low input of negative or unacceptable behaviour. Problematic marriages have a low level of these mutually rewarding behaviours on both sides, or may have a gross imbalance in the input from the two different partners. Instead of exchanging positive behaviour, the partners may use coercive methods to try and force the other to stop doing those things of which they disapprove.

The remedy proposed by behavioural marital therapy is that each partner should state their **complaints**, but that these complaints should then be translated into **wishes** for an alternative way of behaving which is more acceptable, and, as a second stage, into **tasks**. It is very useful to concentrate on practical, domestic issues for these tasks, as these are easily grasped, frequently repeated, and more likely to be remembered than more abstract tasks. In principle the tasks for each partner should be linked and reciprocal, but if this is not possible a 'bank account' approach can be used in which each partner builds up a fund of good behaviour and they work out at the end of a period of time whether it has been mutually acceptable. In moving from complaints to tasks one also moves from past to future, and this is one of the most characteristic features of reciprocity negotiation. The therapist is thus more interested in what will happen next week than in what happened last week or last year.

The way that reciprocity negotiation is used in behavioural–systems couple therapy is a little different from its use in behavioural marital therapy. We will usually have the couple in a decentred position while negotiating, and feel that this helps the process both to be effective, and to translate more successfully to their home setting. We also use it quite briefly at different stages of therapy, rather than as the mainstay of therapy throughout.

The tasks developed for each partner in reciprocity negotiation should be:

1 specific,

2 positive,

3 repeatable,

4 practicable, and

5 acceptable to both partners.

They should also be concerned with everyday activities, rather than once-only events such as arranging an overseas holiday. Sometimes sexual problems can be brought in to the negotiation.

Reciprocity negotiation is a well-tried and effective method of couple therapy in those who accept that they have marital problems. It is also an advantage that the therapist here works in a way which is straightforward and takes an adult-to-adult approach. But it is also, in our setting, a way of assessing whether the couple are ready for this sort of intervention; if not, they can be offered a more systemic input until they are more ready to negotiate.

(b) Communication training

The second strategy in the alternative levels of intervention hierarchy is training in communication. This too is part of the behavioural marital therapy spectrum, but not so exclusively, because work on communication is part of most types of couple therapy. The characteristic feature of the form of communication training used in our setting, however, is that it aims for efficient and clear communication, with positive and constructive requests rather than complaints. Other forms of communication training[25] emphasize other skills such as empathy, reflective listening, and supportive comments. In the present form of communication training these are also issues to be considered, but the main emphasis is on issues such as reducing misunderstandings, ensuring that both partners have an equal say, and helping them both to speak from the 'I' position.

Problems encountered in couple therapy amenable to communication training include:

- lack of empathy
- inability to express emotion
- failing to listen
- monologues with no break for feedback
- one partner the spokesperson and the other silent
- mind-reading (i.e. A knowing better than B what is in B's mind)

- sting in the tail (a positive comment followed by a criticism)
- wandering off the topic
- continual criticism.

In carrying out communication training, the therapist first decentres him- or herself, and asks the couple to converse about a relevant topic. When a problem of communication arises the therapist acts as a 'director' and asks them to discuss the topic in another way. If the problem observed is one of lack of empathy, this may include asking one partner to attend to the emotional state of the other, and perhaps to feed back his or her understanding. If it is of inability to express emotion, the therapist may try to intensify the interaction, pointing out the way in which they are holding back their emotions, and encouraging more expressiveness.

The next three problems are connected: failing to listen, talking in monologues, and the 'spokesperson' problem. Remedies can be decentring, encouraging each partner to speak for him- or herself, stopping the talkative partner (perhaps by asking them to listen to what the other partner has to say), and cutting any monologues short by asking for feedback from the other partner. In dealing with mind-reading one may have to be quite diplomatic, because the process is rather similar to psychotherapeutic interpretation, and some partners may feel that this is a legitimate way of giving insight; however, it should be tactfully blocked, usually by asking the partner whose mind is being 'read' to say whether that is what he or she really thinks.

The 'sting in the tail' is dealt with usually by simply pointing it out, but in some cases it can be neutralized by asking the speaker to restate the idea the opposite way round with the 'sting' first. An example of this is given by a man who said 'I realize you were hurt by what I did, but I had no intention to harm you (i.e. you are being oversensitive)'. He was asked to rephrase it as 'I had no intention to harm you, but I realize that you must have been hurt', and his wife found this much more acceptable, because she could respond to the more positive part of the comment.

The problems of wandering off the topic and continuous criticism are often rather intractable. One way, however, of keeping them to task is to bring them back frequently to the problem first presented, and ask whether they can concentrate on solving it. In the case of mutual criticism, one way of coping is to slow down the interaction so that each partner speaks only after the therapist has intervened to reframe what has just been said.

As with reciprocity negotiation, communication training is used not as a self-contained therapy in itself, but rather as part of a menu of techniques to be chosen according to the problem presented or observed at the time.

(c) Structural moves in session

The main interventions under this heading are raising arguments (or heated discussions) in the session, reversed role play, and 'sculpting'.

There are many couples in which there is a reluctance to enter any sort of conflict. They avoid differences of opinion, and pretend that there is agreement on almost every issue. The more dominant partner, usually more at ease verbally, effortlessly takes the spokesperson role. The other partner is either silent much of the time or spends much effort placating the other in order to reduce conflict.

One strategy with such couples is to ask them to argue (or, to put it more acceptably, to have a **heated discussion**) about a fairly trivial topic. An example of this might be whether the toilet seat should be left up or down after it has been used. It must be a genuine difference of opinion, and not simply one manufactured for the purpose, but it is important that it should be of a trivial nature, as otherwise the couple may feel inhibited about discussing it.

They are then asked to discuss the issue with the therapist observing, and the therapist particularly encourages the more submissive partner to participate with enthusiasm. It may be necessary to ask him or her to speak louder, or to ask the other partner to listen more carefully to what the quiet partner has said, but the therapist should not take sides as such. What is being dealt with is not the issue itself, but the process of arguing. The outcome does not matter, except that the submissive partner should not be allowed to 'get away with' their usual tactic of giving in for the sake of peace. The couple may 'agree to differ' or the submissive partner may have a better than usual hearing, and even win the argument.

This intervention is particularly useful for those couples where there is a degree of depression in the quieter partner, or where the quieter partner is very reluctant to be involved sexually, and is blaming him- or herself.

Another intervention in session which can have an impact on the interaction is the '**reversed role play**'. Here the couple is asked to discuss a particular issue, but they are asked to act as if they were the other partner, even perhaps changing chairs for the purpose. The exercise is useful for some couples who have difficulty understanding each other's point of view, and may promote better mutual understanding.

A third intervention in session is the use of '**sculpting**', in which the partners position themselves and each other wordlessly in a kind of tableau to express some aspects of the relationship. For example, a wife who feels herself excluded from her husband's life may place him looking away from her, while her husband might place the two arm-in-arm and facing the same way. Neither position would represent the objective truth, but each would gain some understanding of the views of the other. The different views could also be the subject for discussion in session or during 'homework'. As with reversed role play, sculpting, with the accompanying 'experiential' insight, can be useful in those couples where there is little understanding of the other's point of view.

(d) Timetables and tasks

These are perhaps the most frequently used of our interventions. They are always given as part of the 'homework' at the end of the session, and may be of a behavioural nature or more systemic. Systems tasks are usually used for behaviour which is thought of as being out of control. Thus, they may be used in a couple where there is a jealous partner: this partner would be asked to raise his or her doubts about the other's fidelity, but only at a specified time each day and for a limited period (e.g. half an hour). If the topic comes up at any other time, they are asked to postpone any discussion till the appointed time. This can be frustrating for the jealous partner, although he or she will perhaps be reassured that the other will give the topic his or her full attention at the set time: but for the other partner it can come as a great relief that the issue of jealousy is at last under some sort of control, even in this simple form.

A timetabled task may be used in other situations, for example, when one partner has a series of complaints which the other is rejecting. The couple can again be asked to discuss the issue only

at certain times and for a limited duration. The advantage of a timetable under these circumstances is that the therapist does not have to adjudicate as to who is right or wrong in the content of the argument, but simply deals with the process of arguing by asking the couple to raise their legitimate complaints at home at an appropriate but limited time.

Another frequently used timetable is the 'talk' timetable, in which a couple who do not communicate very often are asked to set a time each day or evening when they can get together for a discussion about the day's events. In cases where there are difficulties with empathy, it may be useful in addition to ask each partner at the daily talk session to repeat back what the other one has said to reassure the other that they have understood what is meant.

One situation which responds particularly well to timetabling is where the male partner is very keen on sex and the female (while not having a sexual dysfunction as such) is much less enthusiastic. Here the partners are encouraged to reach a compromise on the agreed frequency at which sexual relations might occur, and then they agree on a suitable timetable. The day of the week has to be fixed in advance, since if this is not done the usual arguments will ensue as to whether sex should take place that night, and they are also asked to make the chosen night something special, with perhaps a dinner and the telephone disconnected. If, however, the enthusiastic partner suggests sex on another night, the other can simply remind him that it has been arranged and that they should stick to the arrangement. This remedy may seem somewhat crude, and it is often simply a temporary measure. However, it can be said to have virtually saved some relationships, because it takes the heat out of the sexual conflict which could otherwise lead to divorce, and its use can open up the discussions in subsequent sessions to include non-sexual topics which would otherwise be pushed out by the sexual issue.

It should be mentioned here that many relationship problems have a sexual and a general dimension. When the sexual difficulty is motivational rather than dysfunctional, it is often most productive to deal with it in couple relationship therapy either alone or in combination with psychosexual therapy. In such a case it is quite appropriate to suggest techniques such as the Masters and Johnson 'sensate focusing'[26] in addition to the couple therapy approaches already mentioned.

(e) Paradoxical interventions

Paradox is a relatively infrequently used option in couple therapy (Fig. 6.3.7.1), and is brought in when other methods are ineffective or where the couple relationship seems so rigid that no other intervention can be used. The rationale for paradox depends on a systemic hypothesis which states that the homeostatic forces in the system may be so strong that no straightforward intervention will alter it. All systems tend towards a resistance to change, but in some the resistance is maintained by powerful forces which themselves seem to be informed by extreme anxiety.[14] In these couples or families the only intervention likely to succeed in changing the system is one which prohibits change, but for unacceptable reasons. Although the above explanation is somewhat unsatisfactory, paradox remains in practice a technique which can unlock an otherwise stuck relationship and get the couple back on course for continued therapy.

Paradox is always applied with care and in a sympathetic manner. A common form is to 'prescribe the symptom', that is to advise the couple that it is best 'for the time being' to persist with both the behaviour complained of and the reciprocal behaviour in the other partner. The reason given for this conclusion is a plausible, but challenging and perhaps unacceptable, explanation based on systemic understanding of the relationship.

In using paradox the therapist should think in four stages. First, there should be a positive connotation of the 'symptom' and the reciprocal behaviour. Secondly, there should be a rehearsal of why they are at present helpful for the couple. Thirdly, a statement should be made of the hypothesized feared consequences if the behaviours were to stop. Fourthly, the symptom and the reciprocal behaviour should be prescribed. A case example may make this process a little clearer.

Case Study: A couple who presented with depression in the wife (Edna) and a rather overprotective attitude on the husband, George's, side were in therapy for some weeks without much progress. Following a session in which the therapist asked many questions of both of them about the circumstances and consequences of the depressive episodes, the paradox was presented as follows. 'This depression seems in some ways to be quite good for you as a couple, because it enables Edna to help George by giving him a role in life as her protector. If the depression were to disappear it might be difficult for you both to continue your peaceful relationship, because the differences between your views and ideals would become very clear and you might argue all the time. So for the time being it is better for Edna to remain depressed and for George to be her spokesman and protector'.

This intervention led to quite an outburst from the wife, who up to that time had always been very quiet, and she began to talk of some of the differences of opinion that they had actually had. The husband looked rather disconcerted, and questioned the therapist's reasoning. In the next two sessions the couple reversed their imbalance to some extent, the wife became more assertive than the husband, and her depression became less severe.

The paradox can thus be a powerful mechanism for change, but it must be used with some caution, since an instruction given paradoxically may be taken literally. So it would be inappropriate to include in a paradox any instructions to break the law, to harm oneself or others, or to act irresponsibly. If given as recommended, however, the paradox can unlock a 'stuck' system and put the couple on the road to change and improvement.

In a team setting it is probably best to use a **'split team'** message rather than a paradox as such. This presents the paradox as above, but in the form of an alternative, for example in the form of a disagreement between the therapist and the supervising team. 'I feel that you can carry on with the tasks that I have been giving you, but my team think I am being naïve, and that you really need the depressive symptoms and the overprotection to keep your marriage from falling apart'. The effect is similar, but the impact is softened somewhat by this technique.

Couple therapy with couples from other cultures

As mentioned above, most western countries are now multicultural, and a significant minority of those seeking therapy, especially in urban centres, are from immigrant backgrounds. Probably the most frequently seen in Britain are those of South Asian origin, but Eastern European, African and African- Caribbean

couples are also seen quite often. In the USA there are many people of Latin American and Oriental backgrounds, and here again these will be more commonly encountered in urban settings. The author's experience is mainly with South Asians, and the examples will be mainly from this group.

Cultural factors in counselling have been highlighted by d'Ardenne and Mahtani,[27] and include the need for awareness in the counsellor that he or she also comes from a specific culture which may be just as difficult for the client to comprehend as the client's is to the counsellor. They emphasize the need for humility in the face of difference, and the responsibility of the counsellor to check with the clients before making assumptions about their lifestyles and beliefs. Their advice on the use of interpreters is that unofficial interpreters, including members of the clients' own families, should be discouraged because they are inclined to act as therapists themselves, may translate inaccurately, may ignore cultural differences and may even exploit the clients. It is better to use official interpreters, though this may become expensive to the treatment unit, and the interpreters themselves may translate inaccurately in accord with what they think the therapist wants to hear. Ideally a unit would have multilingual counsellors, but this is not always practicable. In practice, it is often possible with a modicum of understanding of the language for a couple relationship therapist to carry out therapy in English without an interpreter, but with a little bit of necessary translation by the partner whose English is better.

Little has been written specifically on cultural factors in relationship therapy, but Ahmed and Bhugra[28] have reviewed the role of culture in sexual dysfunctions, and their observations are also relevant to couple therapy. They emphasize the need to adapt the techniques of 'western' sex therapy to accommodate the cultural backgrounds of the patients, in particular the gender roles in the culture, and they highlight the risks of ignoring these in therapy.

Couple therapists need to be flexible in regard to the aims of therapy in these couples, which may be rather different from the typical white British couple, and there may also be limits to the kind of therapeutic change possible. For example, a couple from South Asia may be orientated towards a male dominated marital pattern, and both partners may be reluctant to accept the kind of equal relationship that typical couple therapy would expect. In other cases, the more traditional male partner may be concerned to retain the traditional dominance, while the (more westernized) woman may be demanding equality. Similar problems may arise in couples who come from different cultures, and there are clearly more interracial relationships developing as the cultures become more integrated.

Religious considerations may bring difficulties to therapy. Strict Muslim couples may be reluctant to attend therapy together, particularly with a male therapist, because of the difficulty of the wife talking to a male stranger. Masters and Johnson[26] found that one of the most reliable prognostic factors in their therapy was the negative effect of any strong religious belief on the outcome. It has also been observed that in a sexual dysfunction clinic Asian couples were more likely than white couples to default from therapy, a finding put down to their pursuit of organic explanations for the problems and educational and language barriers.[29]

Similarly African-Caribbean couples will have different aims and limitations from white British couples. The father in these families may traditionally be more of an absentee, and leave the upbringing of the children to his partner, who becomes the main authority figure in the family. Again, the therapist must remain aware of possible differences from his/her own culture, and remember that the key consideration is the wishes and wellbeing of the couple rather than any imposed set of rules derived from theory. In particular it is usually impossible to persuade the man to take a more active role in child care, even when the woman wants this, and their relationship will usually remain 'semi-detached'.

One particular 'problem constellation' which the author has seen many times is that of an Englishman married to a North American woman. This should theoretically present no problems, as the language and cultures of both are similar. However, they are divided by the tendency for the British man to be reserved and sometimes resentful and the American woman to be outspoken and critical. Such difficulties have also been found with couples from other disparate backgrounds, such as North American and Latin American partners, and Southern European and British couples. These differences in outlook can lead to repetitive quarrels, in which neither partner can understand where the other is coming from, and often there is also a lack of sexual contact between them. In therapy it is usually necessary to explain each partner to the other, using positive terms and helping them to appreciate the cultural differences without condemning the partner. Then they can usually cooperate with reciprocity negotiation and communication training.

Although the examples given are mainly of South Asian and Western couples, the principles for dealing with cross-cultural relationships of all sorts are basically similar. The therapist needs to respect the differences between their culture and his/her own, trying not to impose solutions which are alien to the couple's own culture. In those cases where they each come from a different background, the general approach is to try to build bridges between them, and use the techniques of behavioural systems therapy to solve their difficulties.

Efficacy of couple therapy

The efficacy of couple therapy is not an easy topic to discuss. Problems arise as to how one should assess efficacy, and while most authors would agree that a measurable improvement in marital adjustment is a valid measure of improvement, some authors dismiss that as being too subjective or too superficial. On the other hand, to use an objective criterion such as divorce as an outcome variable might be seen as being too strict on the therapy, since divorce happens for many reasons, and it might not actually be a bad outcome in some relationships.

A review of efficacy in couple therapy has been carried out by Baucom *et al.*[30] They did a very thorough search of the literature, and made some far-reaching and challenging observations. They comment that the untreated improvement rate is very low in couple problems, and that many of the non-behavioural approaches are of unproven efficacy. They conclude however that behavioural marital therapy (comparing mean effect size over a series of 17 independent controlled outcome studies) is an efficacious and specific intervention for marital distress. The improvement is likely to last for up to a year after treatment but there is less certainty over longer follow-up periods. The addition of cognitive restructuring to behavioural couple therapy did not add anything to the efficacy, but the numbers were rather small.

Snyder and Wills[31] evaluated the outcomes of behavioural versus insight-orientated marital therapy, and found that there was equal improvement in the two conditions, with both being superior to waiting list controls. There was, however, a difference at follow-up, with more of those who had had behavioural therapy divorcing than those who had had insight-orientated therapy.

In a controlled study of behavioural versus interpretative couple therapy, Crowe[32] found that both approaches were effective, but that the behavioural approach produced results more quickly. The follow-up at 18 months showed both methods to be of lasting efficacy, with no differences between them at that point.

Another treatment approach, emotion-focused therapy,[33] has produced good results with couples in therapy. As with behavioural couple therapy the couples improved significantly more than those on a waiting list, but this therapy seems less effective in couples with higher levels of distress.

In one of the relatively few studies on the efficacy of systems-orientated couple therapy. Emmelkamp *et al.*[34] evaluated the effects of behavioural versus systems couple therapy, and concluded that the two approaches had very similar results, but both did better than waiting-list controls.

The London Depression Study[20] (see above) found that systemic couple therapy produced good improvement not only in terms of the couple satisfaction but also on the depression in the depressed partner. The therapy was also associated with a lower drop-out rate than the antidepressant condition, and was thus more acceptable to the patients and their partners.

Thus, the two components of behavioural–systems couple therapy have both been validated by outcome research, although at a much higher level for the behavioural than the systemic. It would be desirable to carry out research on the combined therapy, but this has not yet been done, and the best that can be said is that it is a combination of two probably effective treatment approaches, and therefore likely to be effective.

Training

Training for work with couples using a behavioural–systems approach has been thoroughly reviewed by Crowe and Ridley.[1] It requires an ability in the therapist to understand and use different approaches, and an ability to adapt ones activity to the needs of the couple. Before beginning to work as a couple therapist the trainee should have a basic understanding of the dynamics of couple and family interaction, the phases of human development, the impact on the individual of life events, sexual function and interaction and the impact of physical illness on couple and family relationships. These can be dealt with in the traditional seminar format, in which the trainee can also learn about theoretical and technical aspects of the approach to couple therapy itself. It is also important in selecting candidates for training to ensure that they have some experience of counselling individuals or of being in a therapeutic role, for example as a nurse or doctor.

In addition to the seminars there are also more active training sessions in which the trainee is given the technical skills to carry out therapy. The latter take three main forms: role play, observation, and supervized practice. In role play the trainees are encouraged to use either an existing couple or a fictional one and role play a couple therapy session. Ideally they should each, in different exercises, have the opportunity to play the husband, wife, therapist, and observer in therapy.[35] This helps both in the development of technical skills and in learning to be empathetic to clients through having experienced the client role. In role play the trainee can practise any of the techniques required in therapy, but it is perhaps especially useful in the area of communication training, in which the therapist needs to be alert to the problems shown by the partners and able to apply the appropriate technique smoothly and effectively.

In observation and supervision other aspects of the therapy can be taught, especially the more systemic methods such as arguments, reversed role play, and the framing of messages. The trainees move quite quickly from live observation of therapy in the clinic to being firstly a co-therapist and then the sole therapist in the session, supported and supervised by the trainer and the observation team on the other side of the screen. It is in this activity that trainees begin to display their skills or deficits as therapists, and the trainer must be able to assess progress at this stage and take remedial action if a trainee does not seem to be working as well as expected.

Conclusions

The field of couple therapy is a wide and varied one, and there are almost as many different approaches to treatment as in individual psychotherapy. The relatively brief therapeutic method presented here, behavioural systems couple therapy, is an eclectic one, taking techniques from two approaches of proven efficacy and combining them into a flexible and versatile therapy capable of being used in a wide variety of presenting problems. These include simple relationship problems, psychosexual problems, and such psychiatric conditions as anxiety, depression, and morbid jealousy. It is relatively easy to teach, and although it has not yet been subjected to controlled trials it can be assumed to be no less effective than its component therapies which are both effective. It has recently also been recommended in a package for self-help[36] with homework exercises and theoretical explanations to be used without the intervention of a therapist. There are few contraindications for the therapy, and it can be used both as a therapy in its own right or as an adjunctive therapy in, for example, the treatment of depression, psychosis or sexual dysfunctions. It can thus be a useful addition to the various methods available for the reduction of distress, whether in couples or individuals.

Further information

Crowe, M. and Ridley, J. (2000). *Therapy with Couples: a behaviouralsystems approach to marital and sexual problems* (2nd edn.). Blackwells Science, Oxford.

Clulow, C. (ed.) (2001). *Adult Attachment and Couple Therapy*. Brunner Routledge, London.

Organizations which provide information about how to obtain couple therapy

British Association for Sexual and Relationship Therapy www.basrt.org.uk
British Association for Counselling and Psychotherapy www.bacp.co.uk
United Kingdom Council for Psychotherapy www.psychotherapy.org.uk
Institute of Family Therapy www.instituteoffamilytherapy.org.uk
American Association for Marriage and Family Therapy www.aamft.org

Organizations which provide information on training in Sexual and Relationship Therapy

British Association for Sexual and Relationship Therapy www.basrt.org.uk

Porterbrook Clinic, Sheffield (Sheffield Hallam University) www.porterbrookclinic.org.uk
London South Bank University ww.lsbu.ac.uk/psychology
University of Central Lancashire: Lancashire School of Health and Postgraduate Medicine www.uclan.ac.uk
Relate Institute, Doncaster www.relate.org.uk
American Association for Marriage and Family Therapy www.aamft.org

References

1. Crowe, M. and Ridley, J. (2000). *Therapy with couples: a behavioural–systems approach to marital and sexual problems* (2nd edn). Blackwell Science, Oxford.
2. National Statistics Office (1998). *Population trends.* HMSO, London.
3. Robinson, M. (1991). *Family transformation through divorce and remarriage*. Tavistock Press, London.
4. Daniell, D. (1985). Marital therapy: the psychodynamic approach. In Marital therapy in Britain, Vol. 1 (ed. W. Dryden), pp. 169–94. Harper and Row, London.
5. Clulow, C. (ed.) (2001). *Adult Attachment and Couple Psychotherapy*. Brunner Routledge, London.
6. Bowlby, J. (1969). Attachment and loss, Vols 1 and 2. Hogarth Press, London.
7. Wile, D.B. (1993). *Couples therapy, a nontraditional approach* (2nd edn). Wiley, Chichester.
8. Stuart, R.B. (1980). *Helping couples change*. Guilford Press, New York.
9. Liberman, R.P. (1970). Behavioural approaches in family and couple therapy. *American Journal of Orthopsychiatry*, **40**, 106–18.
10. Beck, A. (1988). *Love is never enough.* Harper and Row, New York.
11. Dryden, W. (1985). Marital therapy, a rational emotive approach. In Marital therapy in Britain, Vol. 1 (ed. W. Dryden), pp. 195–221. Harper and Row, London.
12. Minuchin, S. (1974). *Families and family therapy*. Tavistock Press, London.
13. Haley, J. (1980). *Leaving home.* McGraw Hill, New York.
14. Selvini Palazzoli, M., Boscolo, L., Cecchin, G., *et al.* (1978). *Paradox and counter–paradox.* Aronson, New York.
15. Segraves, R.T. (1982). *Marital therapy: a combined psychodynamic–behavioural approach.* Plenum Medical, New York.
16. Weeks, G.R. (1989). *Treating couples, the intersystem model of the Marriage Council of Philadelphia.* Brunner–Mazel, New York.
17. Spinks, S.H. and Birchler, G.R. (1982). Behavioural–systems marita l therapy: dealing with resistance. *Family Process*, **21**, 169–85.
18. Berg–Cross, L. (1997). *Couples therapy*. Sage, Thousand Oaks, CA.
19. Jacobson, N.S. and Margolin, G. (1979). *Marital therapy: strategies based on social learning and behavioral exchange principles.* Brunner–Mazel, New York.
20. Leff, J., Vearnalls, S., Brewin, C.R., *et al.* (2000) The London depression intervention trial. Randomized controlled trial of antidepressants v. couple therapy in the treatment and maintenance of people with depression living with a partner: clinical outcome and costs. *British Journal of Psychiatry*, **177,** 95–100.
21. De Silva, P. (1997) Jealousy in couple relationships: nature, assessment and therapy. *Behaviour Research and Therapy*, **35,** 937–85.
22. McFarlane, W. (2000) Psychoeducational multi-family groups. Adaptations and outcomes. In *Psychosis: Psychological Approaches and their Effectiveness.* (ed. B. Martindale, A. Bateman, M. Crowe and F. Margison) Gaskell, London.
23. Crowe, M. (2005) Couples and mental illness. *Sexual and Relationship Therapy*, **19**, 309–10.
24. Thibault, J.W. and Kelley, H.H. (1959). *The social psychology of groups.* Wiley, New York.
25. Olson, D.H., McCubbin, H.I., Barnes, H., *et al.* (1983). *Families: what makes them work.* Sage, Los Angeles, CA.
26. Masters, W.M. and Johnson, V.E. (1970). *Human sexual inadequacy.* Little, Brown, Boston, MA.
27. d'Ardenne, P. and Mahtani, A. (1989). *Transcultural Counselling in Action*. Sage, London.
28. Ahmed, K. and Bhugra, D. (2007). The role of culture in sexual dysfunction. In *Psychiatry: Sexual Disorders and Psychosexual Therapy*. (ed M.Crowe). Medicine Publishing (Elsevier), London.
29. Bhui, K. (1998). Psychosexual care in a multi-ethnic society. *Journal of Social Medicine*, **91,** 141–3.
30. Baucom, D.H., Shoham, V., Mueser, K.T., *et al.* (1998). Empirically supported couple and family interventions for marital distress and adult mental health problems. *Journal of Consulting and Clinical Psychology*, **66**, 53–88.
31. Snyder, D.K. and Wills, R.M. (1989). Behavioural versus insight orientated marital therapy: effects on individual and interpersonal functioning. *Journal of Consulting and Clinical Psychology*, **57**, 39–46.
32. Crowe, M.J. (1978). Conjoint marital therapy: a controlled outcome study. *Psychological Medicine*, **8**, 623–36.
33. Johnson, S.M. and Greenberg, L.S. (1985). Differential effects of experiential and problem–solving interventions in resolving marital conflict. *Journal of Consulting and Clinical Psychology*, **53**, 175–84.
34. Emmelkamp, P.M.G., van der Helm, M., MacGillavry, D., *et al.* (1984). Marital therapy with clinically distressed couples: a comparative evaluation of system–theoretic, contingency contracting and communication skill approaches. In *Marital interaction: analysis and modification* (ed. K. Hahlweg and N. Jacobson), pp. 36–52.Guilford Press, New York.
35. van Ments, M. (1983). *The effective use of role–play. A handbook for teachers and trainers.* Kogan Page, London.
36. Crowe, M. (2005) *Overcoming Relationship Problems.* Constable Robinson, London.

6.3.8 Family therapy in the adult psychiatric setting

Sidney Bloch and Edwin Harari

The term 'family therapy' covers a range of approaches. At one extreme, it is a method which seeks to help an individual patient. At the other extreme, the focus is on the relationships between people; according to this view psychopathology reflects recurring, problematic interactive patterns among family members. Midway between the two positions is one that views the family as acting potentially either as a resource or a liability for an identified patient. In this chapter, we cover the spectrum but confine ourselves to the adult psychiatric setting.

A historical and theoretical context

The family has long been recognized as a core aspect of social organization. The folklore of all cultures emphasize the family's role to mould the character of its members. In the past 150 years academic disciplines, such as anthropology and sociology, have studied the various forms of family structure found in different cultures, and at different times. Since the 1960s, psychiatry has also developed a clinical and research interest in the family beyond that of genetics.

Scattered through Freud's writings are interesting comments about marital and family relationships and their possible roles in

both individual normal and abnormal development.[1] His description of unconscious processes like introjection, projection, and identification illuminate how individual experiences may be transmitted across generations. In 1921, J.C. Flugel published the first comprehensive psychoanalytic account of family relationships.[2] Influenced by Anna Freud, Melanie Klein, and Donald Winnicott, the child guidance movement in Britain, mainly consisting of social workers, devised a model of one therapist working with the disturbed child and another with the mother. The two clinicians then collaborated in order to appreciate how the mother's anxieties distorted her perception and handling of her child, leading to developmental difficulties.

Proliferation of theoretical schools

Psychoanalytic and related approaches

Things took a different turn in the United States where Nathan Ackerman[3] began in the 1950s to treat families with a disturbed child, using psychodynamic principles. An interest in working with two or more generations arose concurrently with 'transgenerational'-oriented family analysts using object–relations concepts. Thus, Murray Bowen[4] noted that the capacity of psychotic children to differentiate from their families, while still retaining a sense of age-appropriate belonging, was impaired by the effects of unresolved losses and other trauma in parental and grandparental generations. He also devised the genogram, a schematic depiction of family structure, with a notation for notable events; this remains a standard part of family assessment (see below).

Boszormenyi-Nagy and Spark[5] similarly addressed the transgenerational theme, describing how relationships were organized around a ledger of entitlements and obligations, which conferred on each family member a sense of justice or injustice about their situation. This, in turn, reflected childhood experiences of neglect or sacrifices made on another relative's behalf for which redress was sought in adult life.

Systems-oriented (see later)

Bowen[4] also introduced the principles of 'systems theory' into family therapy. A system is defined as a set of interrelated elements that function as a unity within a particular environment and where the whole is larger than the sum of the parts. 'General systems theory', propounded in the 1940s by a German biologist,[6] contains among its key concepts the place of hierarchy and the emergence of new features in the system as it transforms itself, necessarily, from one level of organization to another. A family is an example of a partially open system that interacts with both its biological and socio-cultural environments and changes over time to accommodate developments such as the advent of a first child or the death of a grandparent.

Working with delinquent youth, Salvador Minuchin recognized the relevance of systems thinking. The youngsters often came from poor, emotionally deprived families, headed by a demoralized single parent (usually the mother) who alternated between excessive discipline and helpless delegation of responsibilities to a child or to her own critical mother. Since these families were beyond the reach of conventional 'talking' therapies, Minuchin applied action-oriented techniques which enabled him to 'join' the family and to re-establish an adaptive hierarchy and effective boundaries between subsystems (marital, parent–child, siblings).

Later, treating 'psychosomatic families' where the problem was a child or adolescent suffering from anorexia nervosa, unstable diabetes or asthma, Minuchin and his colleagues noted that these families, while intact and articulate, were often enmeshed. Members avoided challenging the apparent sense of family unity. Typically, marital conflict was detoured through the symptomatic child, resulting in maladaptive coalitions between parent and child (sometimes between grandparent and child) and the involvement of third parties (e.g. helping agencies) in family life; loss of hierarchy and boundaries ensued. Because words were used to avoid change in these well-educated families, non-verbal strategies were devised to face unspoken fears of conflict and change.[7]

Jay Haley's 'strategic therapy'[8] combined features of Minuchin's model with ideas of Milton Erickson whose techniques had skilfully exploited the notion that a covert message lurks behind explicit communication, which defines the power relationship between family members. Related theoretical developments took place in Palo Alto, California in the 1950s, where a group of clinicians, together with the anthropologist Gregory Bateson,[9] observed that implicit in communication were tacit, non-verbal 'meta-communications' which defined the ties between participants. A contradictory quality between these two levels of communication—in which messages carried persuasive, moral, or coercive force for the recipient—formed part of what they called a 'double-bind'; this form of entrapment was proposed, albeit erroneously, as a possible basis for the formal thought disorder found in schizophrenia.[10,11]

(a) Systems-oriented models: further developments

All the above system-oriented views assume that family functioning can be objectively studied. However, therapists are not value-free and may actively orchestrate changes in accordance with their preferred theoretical model; neglected in these circumstances are therapists' biases and their influence.

This tendency probably reflected the determination of family therapists to distance themselves from psychoanalytical theory; but it also led them to neglect the family's past history and changes through the lifecycle, including the relevance of traumatic events.

In response to this criticism there was a shift away from a problem-focused approach, which had typified most communication-based views of psychopathology. The so-called Milan school[12] (see course of therapy below), whose founders were psychoanalysts, launched profound conceptual changes in how to approach the family, particularly in interviewing them. Another innovation was the participation of observers behind a one-way screen whose task was to offer hypotheses about the family-plus-therapist system to the protagonists.

A Norwegian group[13] took the idea one step further by developing the 'reflecting team dialogue'. Here, following a session, the family could observe the therapeutic team discussing their problems and possible causes, and what factors might have prompted them to seek certain remedies—especially those they had persevered with despite the clear lack of effectiveness.

(b) Post-modern developments

Family therapists also began to ask whether families might be hampered from trying out new ways to solve their difficulties because of the ways they themselves had interpreted their past experiences or unwittingly absorbed the explanatory narratives of external 'experts' or society at large.

This led to a shift from considering the family as a system defined by its organizational structure to a linguistic-based one. According to this view the narrative a family relates about themselves is a means to integrate in specific ways their past experience and its significance. Other 'stories' are excluded from consideration. For instance, when a family with an ill member talk to health professionals, the conversations inevitably revolve around problems (a problem-saturated description). The family ignore times when problems were absent or minimal, or when they were confined to manageable proportions. A different story might be told if they were to examine the factors that could have led, or still lead, to better outcomes than those currently deemed pathological.

Several narrative-based approaches apply these concepts.[14–16] Philosophically, they align themselves with post-modernism, a movement which challenges the idea that there is a fundamental truth or grand theory known only by the expert.

(c) Criticism of systems approaches

Many criticisms have been levelled at systems-based approaches, these include:

- disregard of the subjective experiences of family members
- neglect of the family's history
- inattention to unconscious motives in interpersonal behaviour
- not addressing the issue of unequal power in a family, particularly violence against women and child abuse, and
- ignoring various forms of injustice based on societal attitudes regarding gender, ethnicity, and class.

This critique has led to integrating systems-oriented and psychoanalytic concepts, particularly those derived from object–relations theory.[17–20] Specific disorders such as schizophrenia[21] and anorexia nervosa[22] have been targeted. Another noteworthy variant of integration is Byng-Hall's[23] synthesis of attachment theory, systems-thinking, and a narrative approach.

Another criticism of systems-oriented approaches is minimizing the impact of material reality, such as physical handicap, or biological factors, in the causation of mental illness, as well as sociopolitical phenomena like unemployment, racism, and poverty. These are obviously not merely the result of social constructions or linguistic games and the distress they may inflict on people are potentially considerable.

The 'psycho-educational' approach and 'family crisis intervention' have arisen in the context of the burden that severe mental illness, particularly schizophrenia, places on the family and the potential for members to influence dramatically the course of the condition. This has led to a series of family interventions:

- educating the family about the nature, cause, course, and treatment of schizophrenia
- providing the family with opportunities to discuss their difficulties in caring for the patient, and to devise pertinent strategies
- clarifying the role of conflict, not only about the illness but also about other relational issues
- regularly evaluating the impact of the illness on the family, both individually and collectively
- helping to resolve other conflicts possibly aggravated by the demands of caring for a enduringly ill person.

This type of work may be done with a single family or with several families meeting together, known as Multiple-Family Group Treatment (MFGT). The latter has emerged as a powerful adjunct to conventional individual-based treatment of schizophrenia, bipolar disorder, major depression, obsessive-compulsive disorder, somatization disorder, and an array of chronic medical conditions. Good results have been achieved in reducing the relapse rate, duration, and frequency of hospitalization and in boosting compliance with medication.[24] Family crisis intervention, initially devised for families with a schizophrenic relative but since applied to other clinical states, operates on the premise that deterioration or a request by the family to hospitalize a member may reflect change in a previously stable pattern of family functioning. Convening an emergency meeting with the patient, spouse, and other key family members may help to avoid admission. Social and institutional forces outside the family often contribute to a crisis, and may precipitate a psychotic episode in a vulnerable member. The 'open dialogue' model of family crisis interviewing, developed in Finland, fosters discussion about such forces, using concepts and techniques derived from, *inter alia*, the Milan school, narrative approaches, and psychodynamic thinking; this integrated perspective has much potential.[25]

Indications

A measure of controversy has dogged the issue of what constitutes the indications for family therapy. Pioneering practitioners claimed, somewhat overzealously, that their methods were suited to most conditions. A more balanced view since the mid-1990s encompasses a consensus that considering the systemic context is advantageous in assessing and treating any psychiatric problem. However, it does not follow that family therapy is the treatment of choice (or even indicated).

Family therapy, it should be stressed, does not constitute a unitary approach, with one principal purpose. The diversity of theoretical models we have alluded to above, with their corresponding techniques, should make this obvious. Regrettably, attempts to link indications to specific models have contributed little to the field.

It has also become clear that DSM or ICD diagnoses do not serve well as a basis for determining indications for family interventions. DSM has a minuscule section, the V diagnoses, covering 'relational problems'; these are limited in scope and not elaborated upon.[26] We are only informed that the problem in relating can involve a couple, a parent, and child, siblings, or 'not otherwise specified'. ICD neglects this relational area entirely.

In mapping out indications, we need to avoid blurring family assessment and family therapy. A patient's family may be recruited in order to gain more knowledge about diagnosis and treatment. This does not necessarily lead to family therapy. Indeed, it may point to marital therapy or to long-term supportive therapy. Thus, we need to distinguish carefully between an assessment family interview and family therapy *per se*.

A typology of family psychopathology, which might allow us to differentiate one pattern of dysfunction from another and so map out corresponding interventions, remains elusive. Empirical evidence is inconclusive and clinical consensus lacking. An inherent difficulty is in selecting dimensions of family functioning central to creating a typology.[27] Communication, cohesiveness, adaptability,

boundaries between family members and subgroups, and level of conflict are a few of the contenders offered (see our own classification below).

There are no clear correlates between conventional diagnoses and family type. Efforts to establish links, such as an anorexia nervosa family[28] or a psychosomatic family[29] have not been fruitful. Similarly, investigations into the family and schizophrenia have yielded no durable results.[4,10] Clinicians and researchers have reluctantly accepted that models of effective family-based treatment for mental illness may not necessarily follow an understanding of the apparent causes of a condition in terms of observed disturbances in family relating. This complex matter is helpfully reviewed by Eisler regarding studies of the treatment of anorexia nervosa, but has implications for the entire field.[30]

What follows is our attempt to distil clinical and theoretical contributions.[31] Given the considerable overlap in clinical practice, categories are not mutually exclusive; and a family may require family therapy based on more than one indication. We should stress that family dysfunction is obvious in certain clinical situations and covert in others, often being concealed by a specific member's clinical presentation. Six categories emerge:

1 The problem manifests in explicit family terms and the therapist readily notes the family's dysfunction. For example, a marital conflict dominates, with repercussions for the children; or tension between parents and an adolescent child dislocates family life with everyone ensnared in conflict. In these situations the family is the target of intervention by dint of its clear dysfunctional pattern, and family therapy undoubtedly is the treatment of choice.

2 The family has experienced a disruptive life event which has led to its dysfunction. These events are either predictable or accidental and include, for instance, suicidal death, financial embarrassment, diagnosis of a serious physical illness, and the unexpected departure of a child from home. Any family stability that prevailed previously has been disturbed; the ensuing disequilibrium becomes associated with family dysfunction and/or the development of symptoms in one or more members. Family efforts to rectify the situation may inadvertently aggravate it.

3 Continuing, demanding circumstances in a family are of such a magnitude as to lead to ineffective adjustment. The family's resources may be stretched to the hilt; external sources of support may be scanty or unavailable. Typical situations are chronic physical illness, persistent or recurrent psychiatric illness, and the presence of a frail elderly member.

4 An identified patient may have become symptomatic in the context of a dysfunctional family; symptoms are in fact an expression of that dysfunction. Depression in a mother, an eating problem in a daughter, alcohol misuse in a father, through family assessment, are adjudged to reflect underlying family difficulties.

5 A family member is diagnosed with a conventional condition such as schizophrenia, agoraphobia, obsessive-compulsive disorder, or depression; the complications are the adverse reverberations within the family stemming from that diagnosis. For example, the son with schizophrenia taxes his parents in ways that exceed their 'problem-solving' capacity; an agoraphobic woman insists on the constant company of her husband in activities of daily living; a recurrently depressed mother comes to rely on the support of her eldest daughter. In these circumstances, members begin to respond maladaptively to the diagnosed relative, which paves the way for a deterioration of her condition, manifest as an enduring or relapsing course.

6 Thoroughly disorganized families, buffeted by many problems, are viewed as the principal target of help. This is apposite, even though, for instance, one member abuses drugs, another is prone to violence, and a third manifests antisocial behaviour. Regarding the family as the core dysfunctional unit is the rationale rather than a focus on each member's individual problems.

To reiterate, family therapy may not be the only treatment indicated. Thus, in helping a disturbed family struggling to deal with a schizophrenic member, supportive therapy and medication for the patient are usually as pertinent as any family treatment. Similarly, an indication for family therapy does not negate the possible use of another psychological approach for one or more family members. For instance, an adolescent striving to separate and individuate may benefit from individual therapy following family treatment (or in parallel with it), while his parents may require a separate programme to focus on their sexual relationship.

Contraindications

These are self-evident and therefore mentioned only briefly.

1 The family is unavailable because of geographical dispersal or death.

2 Shared motivation for change is lacking. One or more members may wish to participate, but their chances of benefiting from a family approach are likely to be less than if committing themselves to individual therapy. We need to distinguish here between poor motivation and ambivalence; in the latter, the assessor teases out factors that underlie it and may encourage the family to engage.

3 The level of family disturbance is so severe or long-standing, or both, that a family approach seems futile, according to the best possible clinical judgement. For example, a family that has fought bitterly and incessantly for years is unlikely to engage in the constructive purpose of exploring their patterns of functioning.

4 Family equilibrium is so precarious that the inevitable turbulence[32] arising from family therapy is likely to lead to decompensation of one or more members; for example, a sexually abused adult may do better in individual therapy than by confronting the abusing relative.

5 The patient is too incapacitated to withstand the demands of family therapy. Someone in the midst of a psychotic episode or buffeted by severe melancholia is too affected by the illness to engage in family work.

6 An identified patient acknowledges family factors in the evolution of his problem, but seeks the privacy of individual therapy to explore it, at least initially. For example, a university student struggling to achieve a coherent sense of identity may benefit more from her individual pursuit of self-understanding. Such an approach does not negate an attempt to understand the contribution of family factors to the problem.

Assessment

Family assessment, an extension of individual psychiatric assessment, adds a broader context to the formulation. The range and pace of the enquiry depends on the specifics of the case. Its phases are history from the patient, a provisional formulation concerning the relevance of the family, an interview with one or more members, and a revised formulation. In some cases, it is clear from the outset that the problem resides in the family group, thus rendering the phases below superfluous.

History from the patient

The most effective way to obtain a family history is by constructing a family tree. Apart from showing the structure, it allows relevant information about noteworthy life events and a range of family features to be added. Scrutiny of the tree also provides a source of issues warranting exploration and, eventually, the potential for formulating hypotheses.

Personal details such as age, date of birth and death, occupation, education, and illness are recorded for each member, as well as critical family events (for example, migration, crucial relationship changes, notable losses, and achievements), and the quality of relationships. For an excellent discussion of the family tree—its construction, interpretation, and clinical uses—see McGoldrick and Gerson.[33] (See Fig 6.3.8.1 for genogram conventions.)

Useful principles are to work from the presenting clinical problem to the broader context, from the current situation to its historical origins and evolution, from 'facts' to inferences, and from non-threatening to more sensitive themes.

Questions are best preceded by a statement such as: 'In order to understand your problems better I need to know something of your background and your current situation'. This can be enriched by questions that allude to interactive patterns: 'Who knows about the problem? How does each of them see it? Has anyone else in the family faced similar problems? Who have you found most helpful and least helpful so far? What do they think needs to be done'. Attitudes of family members can be thus explored and light shed on the clinical picture.

The presenting problem and changes in the family

Questions to understand the current context include: 'What has been happening recently in the family? Have there been any changes (e.g. births, deaths, illness, losses). Has your relationship with family members changed? Have relationships in the family altered?'

The wider family context

A broader enquiry flows logically in terms of other family members to be considered, and in the time span of the family's history. Other significant figures, which may include caregivers and professionals, should not be forgotten.

Apart from information about the extended family's structure, questions about their response to major events can be posed: for example, 'How did the family react when your grandmother died? Who took it the hardest? How did migration affect your parents?'

Relationships are explored at all levels, covering those between the patient and other members and between these other members. Conflicted ties are particularly illuminating. Understanding who takes what 'roles' is also useful: 'Who tends to take care of others? Who needs most care? Who tends to be the most sensitive to what is going on in the family?' Asking direct questions about members is informative, but a better strategy is to seek the patient's views about their beliefs and feelings and to look for differences between members: 'What worries your mother most about your problem? What worries your father most?' Several lines of enquiry may reveal differences.

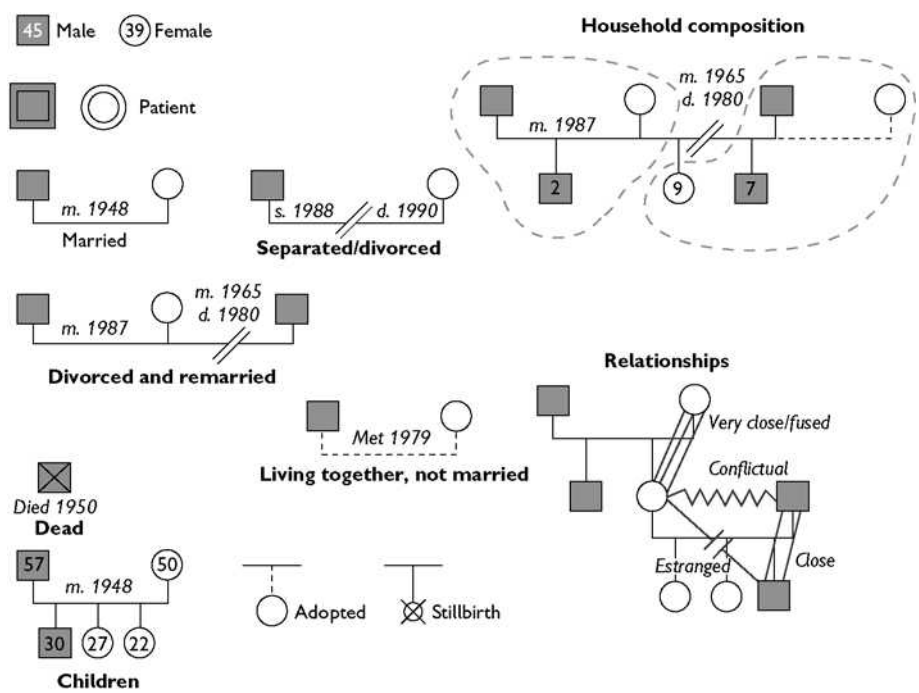


Fig. 6.3.8.1 Genogram conventions.

- Pursuing sequential interactions: 'What does your father do when you say your depressions are dreadful? How does your mother respond when your father advises you to pull up your socks? How do you react when she contradicts him?
- 'Ranking' responses: 'Everyone is worried that you may harm yourself.Who worries most? Who is most likely to do something when you talk about suicide?'
- Looking for relational changes since the problem: 'Does your husband spend more or less time with you since your difficulties began? Has he become closer or more distant from your daughter?'
- Hypothetical questions dealing with imagined situations: 'How do you think your relationship with your wife will change if you don't improve? Who would be most likely to notice that you were getting better?'

Triadic questions help to gain information about relationships which go beyond pairs; for example: 'How do you see your relationship with your mother? How does your father see that relationship? How would your mother react to what you have told me if she were here today?'

Making a provisional formulation

Two questions arise from the above interview—how does the family typically function and are there any family features relevant to the patient's problems?

(a) How does the family typically function?

A schema to organize ideas about family functioning builds from simple to complex observations: structure, changes, relationships, interaction, and the way in which the family works as a whole.

- The family tree will reveal the many family structures possible—single-parented, divorced, remarried, siblings with large age gaps, adoptees. Unusual configurations invite conjecture about inherent difficulties.
- Data will be obtained about notable family changes and events; the timing of predictable transitions is pertinent. Have external events coincided with these transitions (times at which the family may be more vulnerable)? How has the family met such changes?
- Relationships refer to how members interact with one another. What is the degree of closeness and emotional quality (e.g. warm, tense, rivalrous, hostile)? Major conflicts may be noted, as may be overly intense relationships.
- Particular interactive patterns may become apparent which go beyond pairs. Triadic relationships are more revealing about how a family functions overall. A third person is often integral to defining the relationship between another pair. A conflict, for instance, may be re-routed through the third person, preventing direct resolution. A child may act in coalition with one parent against the other or with a grandparent against a parent.
- At a higher level of abstraction, the clinician notes how the family works as a whole. Particular patterns (possibly a series of triads) may emerge that may have recurred across generations. For example, mothers and eldest sons have fused relationships, with fathers excluded, while daughters and mothers-in-law are in conflict.

Idiosyncratic shared beliefs may be discerned, explaining much of the way the family does things. 'Rules' governing members' behaviour towards one another or to the outside world may flow from these beliefs. For example, a family may hold that 'You can only trust your own family; the outside world is always dangerous'; they may therefore avoid conflict at any cost, and prohibit seeking external support.

Evidence of family difficulties may be found at each of these five levels. If they are, the question arises whether these do or do not relate to the patient's problems.

(b) Are family factors involved in the patient's problems?

Links between family functioning and the patient's problems take various forms, but the following categories cover most situations:

- the family as reactive
- the family as a resource, and
- the family in problem maintenance

Often, more than one will apply.

The family as reactive

The patient's illness, or its exacerbation, may have occurred at a time of family upheaval. While the precipitant for the upheaval may have been inherent in the illness itself, an escalating combination of the two may pertain. The illness may have occurred in the face of family stress; it pressurizes the family all the more, and this in turn exacerbates the illness.

The family as a resource

The family may be well placed to assist in treatment. This may be as straightforward as supervising medication, ensuring clinic attendance, and detecting early signs of relapse, or providing a home environment that promotes and maintains recovery. The family may also call on friends and agencies, professional or voluntary, to offer support.

The family in problem maintenance

Interactions revolving around the patient's illness may act to maintain it.

1 First, the illness becomes a way of 'solving' a family problem, the best that can be achieved. For example, anorexia nervosa in a teenager due to attend a distant university may lead to her abandoning this plan since she feels unable to care for herself. Were she to leave, parental conflict would become more exposed and her mother, with whom the patient is in coalition against her father, would find herself unsupported. The illness therefore keeps the patient at home and enmeshed in the parental relationship, and also provides a focus for shared concerns and an ostensible sense of unity.

2 Maintenance of the illness does not solve a family problem but may have done so in the past. An interactive pattern persists even though it lacks utility. In the previous example, the father's mother died 9 months later. His wife subsequently expressed feelings of closeness which he had not experienced for years; their relationship gradually improved. Both parents, however, continued to treat their daughter as incapable of achieving

autonomy, reinforcing her own uncertainty about coping independently if she were to recover.

3 Persistence of illness reflects a perception by the family of themselves and their problems, to which they are bound by the persuasive power of the narrative they have shaped for themselves. This often stems from the health care professional's explanatory schema.

Interview with key informants

The clinician will by now have made an initial assessment of the patient's problems and of the family context. The next step is an interview with one or more informants, usually family members, to corroborate the story, to fill in gaps, to determine influences impinging on the patient, and to recruit others to help. A family meeting is most effective to accomplish these goals.

Implementing the session may prove difficult since the patient may oppose it for all sorts of reasons: symptoms have been kept secret, he regards it as unfair to burden others, he is ashamed of seeing a psychiatrist, he is fearful the family will be blamed, he is suspicious of them, and so forth. These concerns need ventilating, particularly if the family context is pivotal and it is likely that treatment will be enhanced by their involvement. The patient will agree in most cases. Where the safety of the patient or others is threatened, refusal may be overridden on ethical grounds. Otherwise, refusal must be respected. A family session can be suggested again after a more trusting relationship has been established.

Who should be seen depends on the purpose of the interview; generally, all those living in the household are likely to be affected by the patient's illness. The more family factors pertain, the more desirable the attendance by all. The patient's views are sought since he will provide insight into the members he deems crucial to his 'story'.

The family interview

Much information will have been garnered by the time the family is seen. The clinician should consider any biases that may have infiltrated her thinking about the family, and how best to avoid being drawn into alliances. A non-judgemental stance is paramount.

Introductions are made in the initial phase. Names and preferred modes of address are clarified. The clinician then explains the meeting's purpose, details of which may well influence future participation. She invites everyone to share views about the nature and effects of problems they face.

The clinician has an idea about how the patient's problems relate to family function, and can test it out by asking probing questions and observing interactions. This is kept to herself since it is unhelpful for a hypothesis to be offered prematurely. Instead, details about everyday events are sought and inferences drawn later. For example, rather than focusing on 'closeness', questions can be asked about time spent together, whether intimate experiences are shared, who helps with family tasks, and so on.

Triadic relationships can be scrutinized both through questioning (what does A do when B says this to C?) and observation (what does A do when B and C reveal tensions?). The scope for such 'circular' questioning (a method ushered in by the Milan school) is enhanced if several members participate. A third person may be asked to comment on what two others convey to each other when a particular event occurs. This strategy of not posing questions to which the family may have stereotypical responses challenges them to think about their relationships in a fresh way.

Information is elicited that elaborates the family tree. Observations may be made concerning family structure and functioning; for example, who makes decisions, who controls others and in what areas, the quality of specific dyadic relationships, conflict, alliances, how clearly people communicate and how they solve problems. The discussion then extends to all spheres of family life: beliefs, traditions, rules, and values.

Throughout the interview the clinician affirms the experiences of all family members. Concerns are attended to and the members' strengths and efforts acknowledged.

The interview concludes with a summary of what has emerged. The clinician may wish to continue the assessment or recommend family therapy. If the latter, an explanation of its aim and rationale is then given.

Arrangements are made for a follow-up session, purportedly the launch of family therapy *per se*, but in essence a continuation of 'work' in progress.

Revised formulation

Since new information becomes available at each point, the initial formulation is revised as necessary.

Five observational levels —of structure, transitions, relationships, patterns of interaction, and global functioning—are re-examined in terms of the family as reactive, resourceful, or problem-maintaining. We now turn to the course of family therapy.

The course of therapy

With a family approach agreed upon, therapy begins. However, when a family is referred as a group on the premise that the problem is inherently a family one, it is made explicit that the initial stage incorporates assessment.

Given the plethora of 'schools' of family therapy it would be laborious to chart the course of treatment based on each. We shall focus on the Milan approach,[12] but stress that it has undergone refinements. Our account highlights core features but first we comment briefly on the different roles the therapist may assume.

Role of the family therapist

Beels and Ferber,[34] early observers of possible roles for family therapists, divide them into 'conductors' and 'reactors'; this differentiation remains useful since it transcends schools. Virginia Satir[35] is a good illustration of the conductor given that she espoused the notion of family therapist as a teacher who shares her expertise in how to communicate well by setting goals and the direction of treatment. She guided the family to adopt a new form of language to resolve communication problems, which she saw as the root of their troubles. Additionally, the therapist instils confidence, promotes hope for change, and makes them feel comfortable. Conductor-type therapists are explicit authorities, who intervene actively.

The therapist as reactor resonates with, and responds to, what the family exhibits to her. Psychoanalytically oriented and 'pure' systems therapists are representative since they typically share observations about patterns of relating that emerge. We will illustrate this in our account of the Milan school.[12]

The Milan approach

With assessment complete, the therapist (sometimes a pair) meets the family for about an hour. With her preparatory knowledge, she develops a hypothesis about the nature of the family's dysfunction. She has the opportunity on observing patterns *in vivo* to confirm her notions. Patterns usually emerge from the start, making the therapist's job correspondingly easier. Apart from hypothesis-testing, another key task is to engage the family so they will be motivated to reattend. We could interpolate a dictum here: a primary aim of the first session is to facilitate a second session. A vital element in encouraging engagement is for the therapist to promote a sense of curiosity in family members so that they raise questions about themselves and the family as a group.[36]

The chief strategy is circular questioning.[37] Its main purpose is to address the family's issues indirectly; this avoids applying pressure on members and possibly inviting their resistance. For example, the therapist questions an adolescent about how his parents get on with each other, a mother about how her husband relates to the eldest son, a grandmother about which grandchild is closest to the parents, and so forth. This generates illuminating data about individual members and about the family as a group. In this phase, it helps to clarify the hypothesis and to engage participants. It also affords the therapist a greater facility to remain neutral. Because the system and not a patient is the target of change, the therapist is wary of showing any bias.

Various options are then available. If the therapist works as part of a team, her colleagues have busily observed the proceedings through a one-way screen. The family's consent for this will of course have been obtained. During a break, the team—observers and therapist(s)—pool impressions.[38] This is always enlightening since team members note something that others have missed. A consensus evolves, conclusions are drawn and converted into 'messages'. The therapist returns to the family to convey them. The actual messages and their oracular quality comprise a potent intervention, but not necessarily more cogent than circular questions made during the working session. We should mention here that the narrative school has brought with it a de-emphasis on the 'message' on the grounds that 'truth' is a shared construction.

One to three messages are usually given, with maximal clarity. These have a range of purposes including promotion of intersessional 'work'. Homework may be assigned and another session planned (unless termination was set for this point). Meetings commonly occur 3 to 4 weeks apart, and for good reason. During this time, the family, armed with new ideas, tackle them in their day-to-day lives. It is not critical how they go about it but that they do so. As Cecchin has posited,[36] the family's interest in their functioning should have been so aroused that they will be motivated to continue looking at themselves between sessions.

The varied nature of the message makes them classifiable.[39] Messages are supportive, hypothesis-related, or prescriptive. First, the message has a reassuring and encouraging quality and is not related to the hypothesis. A *complimentary message* might be that 'The team were impressed by how open you all were in the session', and a *reassuring message* that 'This is tantamount to a new start for the family and uncertainties are likely'.

Hypothesis-related messages refer to the hypothesis worked out by the team, and may assume diverse forms. It may be stated directly; for example, 'Susan has assumed the role of therapist for her parents and sister to save the family from breaking up'. There may be reference to change such as, 'The team sees John taking responsibility; John and his father's improved relationship has enabled this to occur'. The family may be offered choices related to the hypothesis; for instance, 'The family could risk openness or remain self-preoccupied'. *Paradoxical messages* are a means to communicate a hypothesis which invites the family to revisit a feature of their functioning so that the family's difficulties are positively promoted and explicitly encouraged; for example, 'The team sense that your problem is working for the good of your marriage; sticking with your illness can save the marriage'. More creatively, the paradox may be split, in that the family are told about a divergence of opinion in the team[40]; for instance, some members believe it is too risky for them to communicate openly, others suggest the family can begin to do so. Through a *prescriptive message* the family is given a task. This may or may not be related to the hypothesis. For example, the family is urged to meet on their own before the next session to explore what inhibits a member from relating closely to the others.

Whatever the form of message, the therapist de-emphasizes the pathological status of the patient and applies what the Milan school calls *positive connotation*. This brilliant innovation rests on the premise that all behaviour is purposeful, and that the purpose can be construed positively. An adolescent's 'open grieving' is reframed as sparing the family the anguish of grief. This quality of message calls for creative thinking and flies in the face of the customary view of symptoms as evidence of psychopathology. Again, curiosity enters the picture as the family hears a positive communication concerning an issue they hitherto regarded as abnormal.

The process described continues in succeeding meetings, with attention also paid to family life between sessions. Duration of therapy depends on how entrenched the family dysfunction is rather than on the status of the patient's problems. Thus, systemic change is aimed for, with the family invited to consider a substitute mode of functioning that is feasible and safe. In practice, the number of sessions ranges from 1 to 12. If progress has not been made by about the seventh session, it is likely that alternate ways of helping family and/or patient are needed.

Ending therapy

Termination issues are less profound than in individual or group therapy. The reason is obvious. The family has come as a living unit and will continue as such. Even when the therapist is a pure conductor, the family's intrinsic resources are highlighted so that they can be drawn on after the therapist's departure. Determining the end-point is not usually problematic. There is a shared sense that the work has been accomplished.

A hypothesis (or set of hypotheses) has been introduced, tested, and confirmed. The family system has been examined so that impediments are recognized and understood and better modes of functioning devised and implemented. The family is not required to leave functioning optimally. Instead, termination occurs when there is agreement that the family feels confident to try out newly discovered options.

As alluded to earlier, this may be determined alongside a judgement that the identified patient (or another family member) requires another form of therapy in their own right. An adolescent who has felt unable to separate and individuate is a good example. While family work has explored the system that hindered his

'graduation', the sense prevails that he could benefit further from individual or group therapy. In another example, the parents may conclude, with the therapist's support that they have an agenda which is not pertinent to their children and is therefore best undertaken in couple therapy.

Problems encountered in therapy

Where assessment has been carried out diligently and motivation for change sustained, treatment proceeds smoothly. A crisis may still buffet the group but, rather than being derailed, the family regard it as a challenge with which to grapple.

Family treatment does not always succeed. Indeed, deterioration may occur, albeit in a small percentage of cases. What common difficulties are encountered? The non-engaging family is problematic in that while evidence points to the need for family intervention, members cannot participate in the task, usually because they resist letting go of 'the devil they know'. In another variation, engagement of some members may fail. This is particularly so in the case of fathers who, in the wake of their denial, tend to see the target of therapy as the identified patient rather than the family as a group. This belief may apply to any member.

Missed appointments may punctuate therapy, often linked to turbulent experiences between sessions or apprehension about what a forthcoming session may engender. Like any psychotherapy, drop-out is possible. On occasion, this is appropriate in that the indication for family therapy was misconstrued. In other circumstances, drop-out is tantamount to failure and may stem from such factors as therapist ineptitude, unearthing of family conflict which they cannot tolerate, and inappropriate selection of a family based on faulty assessment.

Given that the family continues as a living group during treatment, they are exposed to all manners of vicissitudes, and these may disrupt the therapeutic work. For example, an overdose by the patient, abrupt marital separation, or admission to a psychiatric hospital may take its toll and undermine treatment.

In discussing termination, we commented on outcome. Not all families will benefit. The family's dysfunction may be so intractable that it proves impervious to change, hypotheses may be 'off the mark', the family may lack sufficient psychological-mindedness, members may retreat in the face of change because of insecurity, and so forth.

Occasionally, dependency becomes a problem as the family discards any vestige of autonomy and only feels secure in the authoritative hands of the therapist. The latter may inadvertently foster such dependency by assuming undue authority, so precluding any sense of a growing partnership. The family's inherent resources are then not permitted expression.

Finally, part of the family may harbour a secret that threatens the principle of open communication. The therapist may be inveigled into a subgroup, although knowing that secrets are not conducive to the therapeutic process. For example, a wife calling the therapist to say she has been having an affair but cannot divulge this to her husband or children lest she hurt them imposes a burden on the therapist and the process (see Bloch *et al.*[27] for an overview of confidentiality in family therapy).

Sound clinical judgement is required in all these situations. Since no ready-made prescriptions are available, the therapist must be aware that difficulties may occur even in a highly motivated, well-selected family. The general principle, however, is to prevent their evolution if at all possible or to recognize them early and 'nip them in the bud'.

Research in family therapy

Selective reviews of the vast research literature in family therapy have been provided by Carr.[41,42] Larner[43] has examined political and conceptual issues raised by family therapists' attempts to conform to the criteria of evidence-based practice while Stratton and his colleagues[44] have looked at the impact of such research on clinicians. An argument for the continuing relevance of single-case studies has been mounted by Datillio[45] and the possibilities of conducting qualitative research in a systems model by Burck.[46]

The long-standing debate between the role of 'common basic factors' versus 'model-specific factors', which has bedevilled research in individual therapy, also pertains to family therapy. Simon[47] has proposed a testable hypothesis, rich in its implications, that the most effective model is one whose concepts and values most closely resemble the world view of the therapist. He offers this as a conceptual bridge between the two approaches to therapeutic research, although he curiously says nothing about the family's world view.

Modifying the psychoeducational model of family-based treatment to incorporate clinically relevant socio-cultural factors has been proposed for the treatment of adolescents and young adults suffering from schizophrenia[48] and depression.[49]

In appraising the contemporary state of family therapy research in adult clinical psychiatry, we may be cautiously optimistic. Immense strides have been made in developing theoretical concepts. As can be seen in the first section of the chapter, we have a rich array of therapeutic approaches.[50] On the other hand, the growth occurred at a dizzy pace, with the inevitable consequence of overload. How can we make sense of so many offerings? Is integration needed to forestall fragmentation? Have we reached the point to pause and reflect? Are we in a position to evaluate the effectiveness of diverse approaches, and for various types of clinical problems?

Observers of the research[51,52] have pointed to a complicating factor in contemplating future work, namely the therapist assembling a natural group, of varying composition, in which the principal goal is to improve its functioning. We face the conundrum of what constitutes optimal outcome and how it is best measured. We can best illustrate this by citing Asen and his colleagues.[53] In their trial of family therapy they had agreed to apply multidimensional measures to assess outcome—at individual, dyadic, and family system levels. At follow-up they noticed changes at the first two levels, but not in the family as a whole. The latter involved ratings of such aspects as communication, boundaries, adaptability, and competence. The researchers were candid in sharing their doubts about how to deal with the divergent findings. Several interpretations were offered: for example, no change was achieved in family functioning, the instrument of family functioning was non-reactive to treatment since it was a trait measure, and an inappropriate model of therapy was applied in the first place. The group concluded that the 'assumptive worlds' of therapists and researchers were under scrutiny rather than the families themselves.

A group in Oxford[54] encountered similar difficulties in their study of consecutive families treated in an adult family therapy clinic. Whereas two-thirds of the identified patients were judged improved at termination, only half the families were rated as

functioning at a better level. Again, the investigators were left with questions of how to determine what had actually been achieved.

A methodologically simpler method is to focus only on the identified patient. The work of Hafner *et al.*[55] exemplifies this choice—a case-controlled evaluation of family therapy in an inpatient setting with subsequent hospital admission data as the chief outcome criterion. Satisfactory as this study is in design, the omission of a family-system outcome measure leaves us ignorant about the level of family functioning following the intervention.

With these tricky matters in mind, what does research need to sort out? The diffuse question of whether family therapy works is of little utility, and is reminiscent of the sterile debate that typified individual psychotherapy outcome research for decades.[56] While subsequent meta-analyses demonstrated that psychological interventions overall exerted useful effects across a range of conditions, the field was still open to the criticism that efficacy of a specific approach for a particular clinical state remained unanswered. Family therapy should not repeat the same mistake; instead of posing the futile question of whether family therapy is effective in adult psychiatry, we should ascertain whether a specific family therapeutic approach, whose character is well identified and measurable, is useful for both the identified patient, with a specific clinical presentation, and his family's functioning, again well defined.

We have good examples of such research. Many intervention studies of families with a schizophrenic member have carefully described the principles of treatment, its rationale, process aspects, and outcome measures in the patient and (in some cases) the family.[57–60]

As mentioned earlier, multiple family group therapy (MFGT) is emerging as the preferred psychosocial intervention for adolescents and young adults with schizophrenia. A comparison with the 'open dialogue' model we mentioned earlier has not been conducted, and may not be feasible, since the latter is part of a comprehensive treatment paradigm.[61] A Danish study[62] which combined MFGT with assertive community-based treatment of young people with schizophrenia reported improvement not only of psychotic features (both positive and negative), but also a decline in use of alcohol and illicit drugs. Given the major problem of co-morbidity, this is a welcome development.

MFGT has also been directed to helping the caregivers of patients with schizophrenia. In a randomized controlled trial, Hazel *et al.*[63] found that those in the treatment programme for over a 2-year period experienced greater relief from distress compared to controls, but they were not able to determine the mechanism for this effect.

Work in the area of affective disorders and family therapy has been innovative, with MFGT, again, gaining popularity.[64] As in schizophrenia, the family treatment aims to reduce members' hostility and criticism expressed toward the depressed patient.

An excellently designed and executed study on anorexia and bulimia nervosa illustrates how outcome research can contribute to the clinical sphere.[65] In a well-controlled study, patients were randomized to either family therapy or 'routine individual supportive therapy', following their discharge from a weight-restoration programme. The family intervention focused on providing the members with information about the eating disorder and the effects of starvation. Parental anxiety was acknowledged and efforts made to help them take control of their daughter's diet. In parallel with improved physical status, therapy turned progressively to typical adolescent issues of autonomy and how these might be achieved. Overall, a structural approach was applied, with systemic and strategic methods added as necessary. Applying these principles to groups of families with adolescents with anorexia nervosa appears to be useful.[66]

While the above studies concerning particular diagnoses, and involving an identified patient, is necessary for progress,[67] this does not preclude outcome studies where the family system is the main target of change. We illustrate this with a particular form of family grief therapy.[68] The model was derived from earlier empirical research on the outcome of family grieving in an oncology setting. A 13-month follow-up yielded five family clusters, two of which were distinctly dysfunctional, two functional, and an intermediate group vulnerable to maladaptive grieving. Three dimensions of family relational functioning were critical: cohesiveness, conflict, and expressiveness. The researchers then developed a treatment model highlighting the goals of promoting cohesiveness, expressiveness, and optimal management of conflict. A screening instrument was found which could readily identify dysfunctional families. A randomized controlled trial was then carried out which showed certain family types benefiting but others remaining unchanged, and a small group even being made worse.[69]

This necessarily schematic account of research on family therapy in adult psychiatry points to action needed in future. We can best summarize what investigators should strive for as: 'Specificity is of the essence'.

Training

From charismatic figures devising innovative methods of family therapy, the field has developed into a worldwide enterprise, with dozens of books, scores of training courses, several journals, and a busy programme of national and international conferences and workshops on offer.[70] Formal training may be given as follows[71]:

1 University-based programmes that regard family therapy as a distinct professional pursuit, with a corresponding corpus of knowledge, and offer degree courses at various academic levels.

2 Free-standing institutes that also see family therapy as a distinct discipline and provide training, generally part-time and of briefer duration than university-based programmes.

3 Within university-affiliated hospitals and clinics that arrange professional training in psychiatry, psychiatric nursing, psychology, social work, and occupational therapy. Although there is a tremendous diversity in the above programmes, most include:

- Supervision of clinical work with the experienced practitioner (and perhaps other students) observing the trainee and family from behind a one-way screen. Some clinicians however consider the one-way screen as dehumanizing. They advocate instead a model of co-therapy between trainee and supervisor, often with other students sitting in the same room as the family.
- Video recording the trainee's work, which she then reviews with the supervisor and fellow students, is widely used. Tapes conducted by eminent therapists are also popular.

Whether training requires familiarity with concepts and techniques of diverse schools or whether it is preferable to develop expertise in only one school remains an open question. The free-standing institutes tend to be run by practitioners of a particular school so

that, after a mostly cursory overview of the field, training concentrates on a specific model. Wendel and his colleagues[72] have proposed a model of training for multidisciplinary mental health settings which places most emphasis on integrating empirically derived knowledge; flexibility is a crucial feature for facilitating its optimal application.

Conclusion

Family therapy has the potential to play a major role in the adult psychiatric setting. As we have commented above, research reveals several promising developments. Clinicians should seriously consider applying this mode of treatment; they will be much rewarded in doing so.

Further information

Luepnitz, D.A. (1988). *The family interpreted: feminist theory in clinical practice*. Basic Books, New York.

Byng-Hall, J. (1995). *Rewriting family scripts. Improvisation and systems change*. Guilford, London.

Kissane, D. and Bloch, S. (2002). *Family focused grief therapy*. Open University Press, Buckingham, UK.

References

1. Sander, F. (1978). Marriage and family in Freud's writings. *Journal of the American Academy of Psychoanalysis*, **6**, 157–74.
2. Flugel, J.C. (1921). *The psychoanalytic study of the family*. Hogarth, London.
3. Ackerman, N.W. (1958). *The psychodynamics of family life*. Basic Books, New York.
4. Bowen, M. (1971, 1981). *Family therapy in clinical practice*. Aronson, New York. (For an incisive critique of Bowen's theoretical contributions, see Miller, R., Anderson, S., and Keala, D. (2004). Is Bowen theory valid? A review of basic research. *Journal of Marital and Family Therapy*, **30**, 453–66.)
5. Boszormenyi-Nagy, I. and Spark, G.M. (1984). *Invisible loyalties: reciprocity in intergenerational family therapy*. Brunner-Mazel, New York.
6. von Bertalanffy, L. (1968). *General systems theory: foundation, development, applications.* Braziller, New York.
7. Minuchin, S. and Fishman, H.C. (1981). *Family therapy techniques*. Harvard University Press, Cambridge, MA.
8. Haley, J. (1976). *Problem-solving therapy*. Jossey-Bass, San Francisco, CA.
9. Bateson, G. (1972). *Steps to an ecology of mind*. Ballantine, New York.
10. Bateson, G., Jackson, D.D., Haley, J., *et al.* (1956). Toward a theory of schizophrenia. *Behavioural Science*, **1**, 251–64.
11. Bateson, G., Jackson, D.D., Haley, J., *et al.* (1962). A note on the double-bind. *Family Process*, **2**, 154–61.
12. Selvini-Palazzoli, M., Boscolo, L., Cecchin, G., *et al.* (1980). Hypothesising-circularity-neutrality: three guidelines for the conductor of the session. *Family Process*, **19**, 3–12.
13. Andersen, T. (1991). *The reflecting team: dialogues and dialogues about dialogues*. Norton, New York.
14. Anderson, H. and Goolishian, H.A. (1988). Human systems as linguistic systems: preliminary and evolving ideas about the implications for clinical theory. *Family Process*, **27**, 371–93.
15. De Shazer, S. (1985). *Keys to solution in brief therapy*. Norton, New York.
16. White, M. and Epston, D. (1990). *Narrative means to therapeutic ends*. Norton, New York.
17. Flaskas, C. and Perlesz, A. (eds.) (1996). *The therapeutic relationship in systemic therapy*. Karnac Books, London.
18. Braverman, S. (1995). The integration of individual and family therapy. *Contemporary Family Therapy*, **17**, 291–305.
19. Cooklin, A. (1979). A psychoanalytic framework for a systemic approach to family therapy. *Journal of Family Therapy*, **1**, 153–65.
20. Luepnitz, D.A. (1988). *The family interpreted: feminist theory in clinical practice*. Basic Books, New York.
21. Ciompi, L. (1988). *The psyche and schizophrenia. The bond between affect and logic*. Harvard University Press, Cambridge, MA.
22. Dare, C. (1997). Chronic eating disorders in therapy: clinical stories using family systems and psychoanalytic approaches. *Journal of Family Therapy*, **19**, 319–51.
23. Byng-Hall, J. (1995). *Rewriting family scripts. Improvisation and systems change*. Guilford, London.
24. McFarlane, W.R. (ed.) (2002). *Multiple family groups in the treatment of severe psychiatric disorders*. Guilford, New York.
25. Seikkula, J. and Olsen, M. (2003). The open dialogue approach to acute psychosis: its poetics and micropolitics. *Family Process*, **42**, 403–18.
26. American Psychiatric Association. (1994). *Diagnostic and statistical classification of diseases and related health problems* (4th edn). American Psychiatric Association, Washington, DC.
27. Bloch, S., Hafner, J., Harari, E., *et al.* (1994). *The family in clinical psychiatry*. Oxford University Press, Oxford.
28. Minuchin, S., Rosman, A., and Baker, L. (1978). *Psychosomatic families: anorexia nervosa in context*. Harvard University Press, Cambridge, MA.
29. Stierlin, H. (1989). The psychosomatic dimension: relational aspects. *Family Systems Medicine*, **7**, 254–63.
30. Eisler, I. (2005). The empirical and theoretical base of family therapy and multiple family day therapy for adolescent anorexia nervosa. *Journal of Family Therapy*, **27**, 104–31.
31. Clarkin, J., Frances, A., and Moodie, J. (1979). Selection criteria for family therapy. *Family Process*, **18**, 391–403.
32. Jenkins, H. (1989). Precipitating crises in families: patterns which connect. *Journal of Family Therapy*, **11**, 99–109.
33. McGoldrick, M. and Gerson, R. (1985). *Genograms in family assessment*. Norton, New York.
34. Beels, C. and Ferber, A. (1969). Family therapy: a view. *Family Process*, **8**, 280–332.
35. Satir, V. (1967). *Conjoint family therapy*. Science and Behaviour Books, Palo Alto, CA.
36. Cecchin, G. (1987). Hypothesizing, circularity, and neutrality revisited: an invitation to curiosity. *Family Process*, **26**, 405–13.
37. Tomm, K. (1987). Interventive questioning: part II. Reflexive questioning as a means to enable self-healing. *Family Process*, **26**, 167–83.
38. Selvini, M. and Selvini Palazzoli, M. (1991). Team consultation: an indispensable tool for the progress of knowledge. Ways of fostering and promoting its creative potential. *Journal of Family Therapy*, **13**, 31–52.
39. Allman, P., Bloch, S., and Sharpe, M. (1992). The end-of-session message in systemic family therapy: a descriptive study. *Journal of Family Therapy*, **14**, 69–85.
40. Papp, P. (1980). The Greek chorus and other techniques of paradoxical therapy. *Family Process*, **19**, 45–58.
41. Carr, A. (2005). Thematic review of family therapy journals in 2004. *Journal of Family Therapy*, **27**, 399–421.
42. Carr, A. (2006). Thematic review of family therapy journals in 2005. *Journal of Family Therapy*, **28**, 420–39.
43. Larner, E. (2004). Family therapy and the politics of evidence. *Journal of Family Therapy*, **26**, 17–39.
44. Stratton, P., McGovern, M., Wetherall, A., *et al.* (2006). Family therapy practitioners researching the reactions of practitioners to an outcome measure. *Australian and New Zealand Journal of Family Therapy*, **27**, 199–207.
45. Datillio, F. (2006). Case-based research in family therapy. *Australian and New Zealand Journal of Family Therapy*, **27**, 208–13.

46. Burck, C. (2005). Comparing qualitative research methodologies for systemic research: the use of grounded theory, discourse analysis and narrative analysis. *Journal of Family Therapy*, **27**, 237–62.
47. Simon, G. (2006). The heart of the matter: a proposal for placing the self of the therapist at the centre of family therapy, research and training. *Family Process*, **45**, 331–44.
48. Weisman, A., Duarte, E., Koneru, V., *et al.* (2006). The development of a culturally-informed, family-focused treatment for schizophrenia. *Family Process*, **45**, 171–86.
49. Breland-Noble, A.M., Bell, C., and Nicolas, G. (2006). Family first: the development of an evidence-based family intervention for increasing participation in psychiatric clinical care and research in depressed African-American adolescents. *Family Process*, **45**, 153–69.
50. Gurman, A. and Kniskern, D. (eds.) (1991). *Handbook of family therapy*, Vol. II (2nd edn). Brunner-Mazel, New York.
51. Gurman, A., Kniskern, D., and Pinsof, W. (1986). Research on marital and family therapy. In *Handbook of psychotherapy and behaviour change* (3rd edn) (eds. S. Garfield and A. Bergin), pp. 565–624. Wiley, New York.
52. Bednar, R., Burlingame, G., and Masters, K. (1988). Systems of family treatment: substance or semantics? *Annual Review of Psychology*, **39**, 401–34.
53. Asen, K., Berkowitz, R., Cooklin, A., *et al.* (1991). Family therapy outcome research: a trial for families, therapists, and researchers. *Family Process*, **30**, 3–20.
54. Bloch, S., Sharpe, M., and Allman, P. (1991). Systemic family therapy in adult psychiatry: a review of 50 families. *The British Journal of Psychiatry*, **159**, 357–64.
55. Hafner, J., MacKenzie, L., and Costain, W. (1990). Family therapy in a psychiatric hospital: a case-controlled evaluation. *Australian and New Zealand Journal of Family Therapy*, **11**, 21–5.
56. Alexander, J., Holtzworth-Munroe, A., and Jameson, P. (1994). The process and outcome of marital and family therapy: research review and evaluation. In *Handbook of psychotherapy and behaviour change* (4th edn) (eds. A. Bergin and S. Garfield), pp. 595–630. Wiley, New York.
57. Falloon, I., Boyd, J., and McGill, C. (1986). *Family care of schizophrenia: a problem-solving approach to the treatment of mental illness*. Guilford, New York.
58. Dixon, L. and Lehman, A. (1995). Family interventions for schizophrenia. *Schizophrenia Bulletin*, **21**, 631–43.
59. Mueser, K. and Bellack, A. (1995). Psychotherapy and schizophrenia. In *Schizophrenia* (eds. S. Hirsch and D. Weinberger), pp. 626–48. Blackwell Science, Oxford.
60. McFarlane, W., Dixon, L., Lukens, E., *et al.* (2003). Family psychoeducation and schizophrenia. A review of the literature. *Journal of Marital and Family Therapy*, **29**, 223–45.
61. Alanen, Y., Lehtinen, V., Lehtinen, K., *et al.* (2000). The Finnish integrated model for early treatment of schizophrenia and related psychosis. In *Psychosis: psychological approaches and their effectiveness* (eds. B. Martindale, A. Bateman, M. Crowe, and F. Marginson), pp. 235–65. Gaskell, London.
62. Peterson, L., Jeppesen, P., Thorup, A., *et al.* (2005). A randomised multicentre trial of integrated versus standard treatment for patients with first-episode psychotic illness. *British Medical Journal*, **331**, 1065–69.
63. Hazel, N., McDonell, M., Short, R., *et al.* (2004). Impact of multiple-family groups for outpatients with schizophrenia on caregivers' distress and resources. *Psychiatric Services*, **55**, 35–41.
64. Keitner, G.I., Drury, L.M., Ryan, C.E., *et al.* (2003). Multiple family group therapy for major depressive disorder. In *Multiple family groups in the treatment of severe psychiatric disorders* (ed. W. Mcfarlane), pp. 244–67. Guilford, New York.
65. Russell, G.F., Szmukler, G., Dare, C., *et al.* (1987). An evaluation of family therapy in anorexia nervosa and bulimia nervosa. *Archives of General Psychiatry*, **44**, 1047–56.
66. Sholz, M., Rix, M., Sholz, K., *et al.* (2005). Multiple family therapy for anorexia nervosa: concepts, experiences and results. *Journal of Family Therapy*, **27**, 132–46.
67. Rowe, C. and Liddle, H. (2003). Substance abuse. *Journal of Marital and Family Therapy*, **29**, 97–120.
68. Kissane, D. and Bloch, S. (2002). *Family focused grief therapy*. Open University Press, Buckingham.
69. Kissane, D., McKenzie, M., Bloch, S., *et al.* (2006). Family focused grief therapy: a randomized controlled trial in palliative care and bereavement. *The American Journal of Psychiatry*, **163**, 1208–18.
70. Liddle, H. (1991). Training and supervision in family therapy: a comprehensive and critical analysis. In *Handbook of family therapy*, Vol. II (2nd edn) (eds. A. Gurman and D. Kniskern), pp. 638–97. Brunner-Mazel, New York.
71. Goldenberg, I. and Goldenberg, H. (1996). *Family therapy: An overview*. Brooks-Cole, Pacific Grove, CA.
72. Wendel, R., Gouze, K., and Lake, M. (2005). Integrative module-based family therapy: a model for training and treatment in a multidisciplinary mental health setting. *Journal of Marital and Family Therapy*, **31**, 357–70.

6.3.9 Therapeutic communities

David Kennard and Rex Haigh

Introduction

Two of the best-known pioneers of therapeutic communities, Tom Main and Maxwell Jones, defined them as follows:

> An attempt to use a hospital not as an organization run by doctors in the interests of their own greater technical efficiency, but as a community with the immediate aim of full participation of all its members in its daily life and the eventual aim of the resocialization of the neurotic individual for life in ordinary society.[1]

> What distinguishes a therapeutic community from other comparable treatment centres is the way in which the institution's total resources, staff, patients, and their relatives, are self-consciously pooled in furthering treatment. That implies, above all, a change in the usual status of patients.[2]

Today therapeutic communities can be defined by a number of common features, but a word of warning. For reasons of historical coincidence, the term is used in the fields of mental health and addictions to refer to two somewhat different treatment models. In the addiction field they are also known as hierarchical, drug-free or concept-based therapeutic communities, or simply addiction therapeutic communities,[3] in contrast to the more democratized programmes in mental health. The two models have similar goals but their methods differ, although there are signs of increasing rapprochement between them. This chapter deals mainly with therapeutic communities in mental health, but reference will also be made to addiction therapeutic communities and those in long-term care settings. It is worth noting that those admitted to a therapeutic community for treatment are usually referred to as residents, clients, or members, rather than as patients.

Defining beliefs

Certain beliefs about human relationships and the nature of therapy are central to therapeutic communities.

1 Staff are not completely 'well' and residents are not completely 'sick'. There is a basic equality as human beings between staff and residents, who share many of the same psychological processes and experiences.

2 Whatever the symptoms or behaviour problems, the individual's difficulties are primarily in his or her relationships with other people.

3 Therapy is essentially a learning process, both in the sense of learning new skills—how to relate to others or deal more appropriately with distress—and learning to understand oneself and others.

Defining principles

A study of one of the best-known therapeutic communities, Henderson Hospital,[4] identified four principles or 'themes' that have come to be widely associated with therapeutic community treatment.

Four principles of therapeutic community treatment

- **Democratization** Every member of the community should share equally in the exercise of power in decision-making about community affairs.
- **Permissiveness** All members should tolerate from one another a wide degree of behaviour that might be distressing or seem deviant by ordinary standards.
- **Communalism** There should be tight-knit intimate sets of relationships, with sharing of amenities (dining room etc.), use of first names, and free communication.
- **Reality confrontation** Residents should be continuously presented with interpretations of their behaviour as it is seen by others in order to counteract their tendency to distort, deny, or withdraw from their difficulties in getting on with others.

Defining aspects of current practice

The generalizability of these principles to newer therapeutic communities is now being questioned and others are developing theoretical frameworks for different therapeutic communities.[5] In 2002, a quality network including most British therapeutic communities started, the 'Community of Communities', with the explicit aim of defining good practice and improving it. In 2006 the first version of 'Core Standards' was published.[6] This comprised 16 standards which were derived from consensus and consultation exercises to determine what practitioners and service users thought reflected the underlying values of therapeutic communities.

Box 6.3.9.1 illustrates a sample of eight of the standards. Note that 'all community members' should be taken to include both resident or client members, and staff.

Background

Evolution of different types of therapeutic community

Communities providing sanctuary for mentally ill people have been known as far back as the fourteenth century at Geel in Belgium. In 1796 the Retreat was opened by the Quakers in York, England, where personal relationships and social expectations in a family-like atmosphere enabled previously dangerous and unpredictable individuals to control and modify their behaviour.[7] This model, known as 'moral treatment', strongly influenced the creation of asylums in Britain and the United States in the first half of the nineteenth century. In the early twentieth century, pioneers in therapeutic education, inspired by a Christian belief in the therapeutic power of love and by Freud's new method of psychoanalysis (see Chapter 3.1), created residential schools for maladjusted children that demonstrated most of the practices and attitudes outlined above.[8] The modern equivalent of communities such as Geel can be found in the intentional communities run by third sector (voluntary) organizations such as l'Arche and the Camphill communities for people with learning disabilities. (The term 'intentional community' avoids language that implies clinical responsibility or a focus on therapy or change, and has been defined as 'a relatively small group of people who have created a whole way of life for the attainment of a certain set of goals'.) A number of therapeutic communities for children and young people now exist as voluntary organizations in the educational sector, as progressive schools, and as long-term treatment units for very disturbed children.

The history of mental health therapeutic communities for adults began during the Second World War, when the psychoanalyst Wilfred Bion was put in charge of the training wing at Northfield Military Hospital in Birmingham, England. His brief attempt in 1943 to establish a therapeutic community failed, but was soon followed by others who were more successful: Tom Main, S. H. Foulkes, and Harold Bridger at Northfield, and Maxwell Jones at Mill Hill Hospital, London. In dealing with psychiatric casualties among soldiers they developed a radical new approach, which was first described in a series of papers in 1946. One of these coined the term 'therapeutic community'.[1] Main and Jones continued to develop different versions of this new method after the war, Main as director of the Cassel Hospital and Jones at Belmont Hospital Industrial Neurosis Unit, which was renamed the Henderson

Box 6.3.9.1 Core standards for therapeutic communities

1 The whole community meets regularly

2 All community members work alongside each other on day-to-day tasks

3 All community members share meals together

4 All community members can discuss any aspects of life within the community

5 All community members create an emotionally safe environment for the work of the community

6 All community members participate in the process of a new client member joining the community

7 There is an understanding and tolerance of disturbed behaviour and emotional expression

8 Positive risk taking is seen as an essential part of the process of change

Hospital in 1958. The Cassel Hospital continues as an inpatient psychotherapy hospital, and Henderson Hospital replicated itself in 2000 to serve national needs for 'severe personality disorder' provision by founding Main House in Birmingham and Webb House in Crewe.

The creation of the National Health Service in 1948 provided the stimulus to address the major problems of institutionalization revealed in a number of studies of large mental hospitals in the United Kingdom and United States.[9,10] In the 1950s and 1960s social psychiatry was in the ascendancy and a number of these hospitals developed what Clark called the 'therapeutic community approach".[11] In the 1970s and 1980s concepts of collective responsibility fell from favour and individualism prevailed, with a decline in the fortunes of therapeutic communities. The 1990s and 2000s have seen a revival of interest in therapeutic communities within more specific mental health contexts, including prisons, personality disorder services, and for the management of people with enduring mental illness in the community. The problem of degraded and poorly functioning inpatient units is now being addressed by attention to establishing and maintaining 'therapeutic environments' in acute settings, in a direct parallel to the 'therapeutic community approach' 40 years earlier.[12,13]

Alongside these developments two other types of therapeutic community have emerged. In 1958 a self-help organization in the United States called Synanon became the prototype for concept-based therapeutic communities for ex-addicts. Phoenix House and Daytop were two major programmes that grew from this, and today therapeutic communities modelled on them can be found in more than 50 countries worldwide.[3] A development that grew out of the antipsychiatry movement in the 1960s is known at Soteria. These are small low-stress family-like environments where psychosis is responded to with intensive therapeutic support rather than medication. These communities are mainly found in Europe.[14]

Scientific background

Therapeutic communities have drawn on the concepts of psychoanalysis, group analysis (see Chapter 6.3.6), humanistic and integrative psychotherapies, and on sociological studies of mental hospitals which identified phenomena such as the total institution[10] and patterns of behaviour associated with psychiatric treatment in institutions.[15] They are also underpinned by studies of the impact of unconscious processes in organizations,[16,17] and by anthropological studies such as that of Rapoport[4] which found a typical pattern of oscillation in the therapeutic community.

A developmental model based on the 'required emotional experiences' of attachment, containment, communication, inclusion, and agency has been proposed by Haigh.[18] This identifies ways in which a range of psychological theories and approaches are relevant to therapeutic community practice, and illustrates how they are replicated in the structures and culture of a therapeutic community. It also proposes that disturbance of 'primary emotional development' (which all humans undergo early in life) can to some extent be made good by a satisfactory experience of 'secondary emotional development' in a therapeutic community.

Technique—how change is brought about

Since the therapeutic community *is* the treatment, managing treatment involves attention to two parallel processes: the progress of each resident through the community, and the effective functioning of the therapeutic community as a whole. Responsibility for managing these two processes ultimately belongs to the staff, though it is shared with the residents when the community is functioning well.

Most if not all the treatment in a therapeutic community takes place in groups and in the everyday life of the community, although some also use individual psychotherapy. The essence of the therapeutic community technique has been encapsulated in two phrases.

1 A *living–learning* situation: this refers to the fact that everything that happens between members of a therapeutic community in the course of living together, and in particular when a crisis occurs, is used as a learning opportunity.[19]

2 *Culture of enquiry*: this refers to the creation not just of certain structures but of a basic culture among the staff of 'honest enquiry into difficulty'. There is a conscious effort to identify and challenge dogmatic assertions or accepted wisdoms.[20]

The basic mechanism of change is not difficult to explain. The therapeutic community provides a wide range of lifelike situations in which the difficulties a member has experienced in their relationships with others outside are re-experienced, with regular opportunities in small group and community meetings to examine and learn from these difficulties. If the therapeutic community is to work as a therapeutic method, all its constituent parts described in this chapter must be in good-working order. This requires a process by which new members adopt the values of the community, emphasizing openness, responsibility, and active participation, and in turn pass these on when the next new members arrive. To operate this mechanism requires both staff and residents to fulfil a number of roles (Box 6.3.9.2).

Staff training

The need for specialized training for leading or working as a staff member in a therapeutic community has been a matter of some debate.[21] There is an argument that the emphasis on egalitarianism and democratization means that this form of treatment is best delivered by people without special training who can just 'be themselves'. Unqualified social therapists often form a key part of the staff complement. However, the other argument is now increasingly accepted, that while being oneself is an important part of the staff role (see above), therapeutic community work requires a high level of skill and knowledge in a number of areas, together with a well-developed capacity for open honest discussion and reflection.

Training courses exist in the United Kingdom, Finland, Norway, the Netherlands, and Greece, but there is no set standard or curriculum. Some relevant theoretical training is obtainable as part of group therapy and systemic therapy courses, and the case has also been made for a placement in a therapeutic community to be part of professional training in psychiatry and psychotherapy. The benefits include learning at first hand about the treatment of personality disorders, experience of group-based treatment, and working as a member of a multi-disciplinary team, thus gaining first-hand experience of institutional dynamics and the way individuals react to their wider social networks. One of the most popular short courses in the United Kingdom and the Netherlands

Box 6.3.9.2 Staff and client/resident roles in a therapeutic community

Role	Role activity of client/resident members	Role activity of staff members
Participation and involvement in the daily life of the community('Living–learning')	To explain how the community works to those referred, to visitors, and new members	To spend informal time with client/residents (those who appear aloof may be challenged about this)
	To take responsibility for various tasks which contribute to the running of the community	To work alongside clients/residents in day-to-day tasks
	To contribute to various 'extras' such as being involved in teaching and research	To monitor other staff members' emotional involvement and consider in supervision
	To notice and include those who isolate themselves	
Contributing to and managing therapeutic processes ('Culture of enquiry')	To use identification to challenge or support peers	To use therapeutic interventions in groups of various modalities (e.g. group analytic, psychodrama, art therapy, CBT)
	To be open to the challenge and support of one's peers	To encourage client/resident members to take a therapeutic role, in some cases delegating responsibility for out-of- hours support to them
	To support those in crisis, often including out-of-hours	
	To maintain community structures such as rules and timekeeping	
Responsibility for decision-making and 'Democratization'	To make decisions about day-to-day and domestic matters, including rotas and elections	To decide the level of decision-making which would be optimal for the therapeutic benefit of the client group, according to their capabilities and needs
	To participate in decisions about therapeutic matters (e.g. deciding consequences of breaking rules)	To monitor the functioning of the therapeutic environment and titrate the level of staff input and leadership required
	To be involved in planning local events and service developments (e.g. open days, starting new groups)	To ensure good communication with managers, commissioners, referrers, clinical colleagues, and other relevant organizations
	Senior and ex-members can offer invaluable support in maintaining effective external relations (e.g. with organizational executives)	

is a brief residential simulated therapeutic community.[22] Here, 20 to 30 health professionals live together for a few days in the roles of residents with a 'staff' group working with them. This provides a valuable opportunity to experience at first hand the workings and impact of this form of treatment.

Indications and contraindications

Universal indications for therapeutic community treatment are difficult to give. Modified therapeutic communities have been developed for people with different types and levels of psychiatric disorder, and even the same therapeutic community may fluctuate in its capacity to absorb difficult members. An individual's suitability will need to be judged in relation to a particular therapeutic community at a particular time. Having made this caveat, the general indications and contraindications will usually apply (Box 6.3.9.3).

Pathways and process: phases in the therapeutic journey

There are usually four distinct but overlapping phases in a member's journey to and through a therapeutic community:

Engagement phase: Referral, preparation, and selection procedures are an integral part of therapeutic community practice, involving both the prospective member and existing residents as active participants in the process, which start with referral or self-referral. Many prospective members of therapeutic communities are wary or fearful of the forthcoming therapy, and need support and encouragement to persist. This is often effectively delivered by current or ex-members, arranged in partnership with voluntary agencies, or with internet support groups. During this time, any regular support from mental health teams or other agencies should continue.

Assessment and preparation phase: When a decision to proceed towards formal treatment has been made, several arrangements may need to be made. These include formal assessment processes, practical planning, and agreeing a treatment contract. This work is frequently arranged through the use of an 'assessment and preparation group', which is also designed to be a time-limited foretaste of what the treatment phase entails. The assessment process can be undertaken in this group itself, in smaller groups, or with individual appointments. The practical planning involves matters such as arranging childcare, securing stable accommodation, and agreeing plans for medication and risk management. It also includes an

Box 6.3.9.3 Indications and contraindications for therapeutic community treatment

Indications	Indications for specialized TC	Contraindications
Diagnosis of:		
◆ Personality disorder	First episode psychosis	Physically dependent addictions
◆ Self-harm	Serious and enduring mental illness	Current mania
◆ Adjustment disorders	Less than 18 years old	Depression with severe retardation
◆ Recurrent depressive disorders	Learning disabilities	Dementia
◆ Bipolar disorder	Perpetrators of sexual abuse	Dangerously low weight
◆ Intractable anxiety disorders		Antisocial PD with history of intimidation and deception
◆ Eating disorder		No capacity for social involvement
◆ Addictions		Inability to see problems in terms of relationships
		Unwillingness to engage in informal, intimate, and open style of relating with professionals
Age:		
No age limits: young children to elderly members can show benefit		Belief that only experts can help

explicit treatment contract, which may be a verbal agreement about understanding the community rules, or a formal written and signed agreement.

Treatment phase: This usually begins with a formal 'case conference' or 'selection panel' including current community members with a decision made by voting. Subsequent therapy programmes vary considerably: from 1 day per week to whole-time residential; from predominantly sociotherapy to a range of psychoanalytic, cognitive behavioural, humanistic and interpersonal and systemic groups; from a few weeks duration to several years, either time-limited or open-ended; and from group size of less than 6 to more than 50. Some communities include individual therapy, but others consider this inimical to the group dynamic process. A modal or typical programme would be for between 3 and 5 days per week, with all interventions in groups comprising a mixture of community meetings, small therapy groups, shared lunch, and informal time together. A typical community would have between 12 and 24 members divided into three small groups, who would stay for 12 to 18 months. Suitable arrangements would be in place for crisis meetings to be called at short notice, as would a system for members to support each other out-of-hours.

During the first few weeks of treatment the new member will be feeling his or her way, forming attachments to one or two others, but still wary of the groups. After the first month or two he or she will begin participating more actively in the groups, taking part in the full life of the community with certain role responsibilities, helping and supporting other members. This will probably include experience of situations similar to those triggering referral, such as having to deal with authority, fear of failure, feeling rejected or abandoned, situations evoking rivalry and competition, or many others. As before, these may trigger destructive or violent impulses towards the self or another person, or the experience of other symptoms of distress. Through the group meetings the member is confronted with the effects of their behaviour on fellow members and the meaning of the behaviour or symptoms is explored, making full use of the insights and understanding of fellow members. Through this repeated process the member gradually comes to experience himself and others differently. As one member wrote: *Bit by bit, almost grudgingly, the fact dawned on me that I wasn't surrounded by forty sticks of furniture but by Jim, Gary, Jane . . .* [23]

Re-entry phase: Until recently, many therapeutic communities had a 'cliff-edge' ending, where one day members are able to have the community's full emotional and practical support, and the next are not allowed to contact any other members. Although this has some theoretical justification in terms of 'coming to term with endings', and has strong advocates amongst ex-members of therapeutic communities, it is now generally considered better practice to support members over the leaving process, and then into re-establishing mainstream social networks. This can be done with a specific 'leavers group', that members join while in the full treatment phase and continue afterwards, for either a fixed or indefinite period. These groups normally include a practical focus, and are social and supportive rather than exploratory and therapeutic. For those who are ready and able, they can have objectives of securing employment or education for members. In the case of prison-based therapeutic communities for ex-addicts the provision of drug-free housing and vocational training have been found to improve success rates.[24]

As well as planned endings there are various other types of ending. Some members may leave prematurely, unable to cope with therapy; some may be 'voted out' by the community for a serious or repeated transgression of community rules. Such endings do not necessarily indicate a treatment failure, although the longer members remain the more likely they are to benefit.

Research evidence

The effectiveness of therapeutic communities has been investigated in relation to different clinical problems, which are discussed separately.

Personality disorders

Until recently there has been little systematic evidence of the efficacy of therapeutic communities for treating personality disorders, and disagreement over whether those who did benefit were really suffering from psychopathic or personality disorder. While efficacy in this area is still questioned (see Chapter 4.12.7), the picture has recently become clearer with the publication of the first systematic

review of therapeutic community treatment for people with personality disorders.[25] The authors carried out a full search of therapeutic community publications and grey literature, collecting over 8000 references from 38 countries. These were reduced to 29 research studies that met the criteria of randomized controlled trial design (eight studies) or comparative or controlled studies that reported raw data and used conservative outcome criteria (e.g. reconviction rates rather than psychological improvement). A meta-analysis found that 19 studies showed a positive effect within the 95 per cent level of confidence while the remaining 10 straddled the neutral score. The overall summary log odds ratio was −0.567, with a 95 per cent confidence interval, −0.524 to −0.614. The authors concluded that there is strong evidence for the effectiveness of therapeutic communities. A more recent systematic review assessed evidence for interventions for people with PD in general and for dangerous and severe personality disordered offenders and made clear recommendations about the most promising treatment interventions for PD in use or currently in development. The reviewers covered therapeutic community programmes; cognitive, behavioural, cognitive behavioural, and psychodynamic psychotherapies; pharmacological and physical treatments. They concluded that 'the TC model currently has the most promising evidence base in this poor field'.[26]

Offending behaviour

Therapeutic communities have been established in prisons to deal with disruptive, violent inmates, and also with the underlying problems of antisocial personality disorder. Results at Barlinnie Special Unit in Scotland demonstrated substantial reductions in violent incidents within prison. Reconviction studies carried out at Grendon, a prison run entirely on therapeutic community lines, found that prisoners had lower rates of reconviction, fewer custodial sentences, and fewer reconvictions for violent offences than prisoners on the Grendon waiting list who never went there. Those who stayed at Grendon longer than 18 months showed the greatest reductions in reconviction rates. Re-offending rates have also been found to be lower in the former Federal Republic of Germany for prisoners in Social Therapeutic Institutions than those receiving standard prison sentences.[27] In 2004, the 'Democratic Therapeutic Community Core Model' was accredited as a treatment programme for use in prisons in England and Wales. Such programmes must show evidence of their capacity to impact on dynamic risk factors known to be associated with re-offending. Risk factors that therapeutic communities have been found to have a positive influence on include negative attitudes towards authority, identification with antisocial role models, and acceptance of responsibility for offending behaviour.

Drug dependence

Therapeutic communities for drug dependence use the hierarchical or concept-based model. These form one part of the range of treatments for drug abuse, which includes other residential models such as Christian communities and the Minnesota model, as well as methadone maintenance programmes and psychotherapy. Therapeutic communities for drug dependence, often modified to suit different local cultures, can now be found in many European and international countries as well as in the United States where they originated. Although the model began, and continues, as a residential peer-support programme in the community, it has adapted well to secure environments, and concept-based therapeutic communities for drug dependence have been established in American prisons since the mid-1980s, accompanied by a growing number of aftercare programmes providing employment and drug-free accommodation. Several national studies have evaluated the outcome of these programmes. Randomized controlled trials show that no-treatment groups have a higher level of recidivism than those who complete treatment in a prison therapeutic community, and that recidivism is further reduced by participation in a community aftercare therapeutic community.[24] Since 1995 a number of these therapeutic communities have also been established in English prisons. The general conclusion, for both secure and non-secure therapeutic communities, is that residents who stay in programmes for longer periods have lower rates of drug use and criminal behaviour and higher rates of employment than those who stay for shorter times. However, there are no direct comparisons between therapeutic communities and other treatment models, and it is likely that therapeutic communities are successful for those who are well motivated. Although this may be only a relatively small proportion of all drug abusers, studies, and admission policies, suggest those who enter these high intervention therapeutic communities tend to be severely addicted and damaged, often dually diagnosed with personality disorder, and less likely to respond to low intervention treatments.[28]

First episode psychosis

A small number of therapeutic communities have been developed in the United Kingdom, United States, Switzerland, and Germany on the principle that first episodes of psychosis can be effectively treated in low-stress family-like settings providing round the clock personal support, with no or minimal use of neuroleptics. This has become known as the Soteria model. Two Soteria houses, in California and Berne, have been subjected to randomized or matched control trials comparing them with usual hospital treatment. One study found that completing subjects with schizophrenia exhibited a large effect size benefit with Soteria treatment, especially in the areas of psychopathology, work, and social functioning. Length of stay in Soteria Berne was initially longer but this was subsequently reduced to less than the admission ward. In both studies the 2-year outcomes were at least as good in the Soteria group and less antipsychotics were prescribed for the Soteria group.[14] A 20-year study of an acute psychiatric ward in Finland found that people with acute psychotic episodes and borderline conditions seemed to benefit from the therapeutic community model with a high level of support, negotiation, order, and organization.

Severe and enduring mental illness

The therapeutic community approach[11] has been widely used in large mental hospitals to counter the effects of institutionalization and to mobilize the residual capacity of those suffering from chronic mental illness for social relationships, purposeful employment, and personal responsibility. The method was as much about improving the sense of purpose and morale of the staff and the general quality of life in the institutions as it was about clinical improvement, and its success was demonstrated in the way some large old mental hospitals were turned into centres of excellence.

With the re-provision of services for people with enduring mental illness in the community, therapeutic community principles have been found to be an effective way of structuring staffed hostels and homes. In one version of this, the 'ward in a house', the model is close to the original practice of the York Retreat, an antecedent of therapeutic communities (see above).[29] The challenges presented by the severely mentally ill chemical user have also been addressed using a modified therapeutic community with some evidence of success. Three main modifications required to the TC structure were increased flexibility, decreased intensity, and greater individualization.[30]

Children and adolescents

Therapeutic communities for children and adolescents were first developed in the field of therapeutic education almost a century ago, and now exist for a variety of needs: learning disability, delinquency, and emotional disturbance. Little systematic evaluation has been carried out. A survey of 186 children in nine therapeutic communities for emotionally disturbed children found evidence of increased stability and hopeful outcomes for those who stayed. A 20-year follow-up of 28 children in one community reported evidence of long-term improvement.[31]

Wright and Richardson, reviewing the current state of research for therapeutic communities for children and young people, conclude that, 'when rigorous quality controls are introduced there are as yet too few studies to draw any aggregated conclusions. Perhaps the clearest qualified statement would be that there is low-level evidence that some residential therapeutic placements produce changes in the mental and social functioning of some young people who have been unable to cope with family life'.[32]

Learning disabilities

Long-term residential communities for adults with learning disabilities exist in the charitable sector in many countries around the world. These are value based rather than evidence based. There is some evidence that many families express a strong preference for village-style communities such as the Camphill for their mentally handicapped relatives.[33]

Further information

Campling, P. and Haigh, R. (eds.) (1999). *Therapeutic communities: past, present and future*. Jessica Kingsley, London.

Campling, P., Davies, S., and Farquharson, G. (2004). *From toxic institutions to therapeutic environments*. Gaskell, London.

Community of Communities. (2006). *Service standards for therapeutic communities*. Royal College of Psychiatrists, London.

Kennard, D. (1998). *An introduction to therapeutic communities* (2nd edn). Jessica Kingsley, London. www.therapeuticcommunities.org—website of the Association of Therapeutic Communities.

References

1. Main, T. (1946). The hospital as a therapeutic institution. *Bulletin of the Menninger Clinic*, **10**, 66–70.
2. Jones, M. (1968). *Social psychiatry in practice*, pp. 85–6, Penguin, Harmondsworth.
3. Rawlings, B. and Yates, R. (2001). *Therapeutic communities for the treatment of drug users*. Jessica Kingsley, London.
4. Rapoport, R.N. (1960). *Community as doctor*. Tavistock, London.
5. Campling, P. and Haigh, R. (eds.) (1999). *Therapeutic communities: past, present and future*. Jessica Kingsley, London.
6. Community of Communities. (2006). *Service standards for therapeutic communities*. Royal College of Psychiatrists, London.
7. Tuke, S. (1813). *Description of the retreat*. Reprinted by Process Press, London, 1996.
8. Bridgeland, M. (1971). *Pioneer work with maladjusted children*. Staples Press, London.
9. Barton, R. (1959). *Institutional neurosis*. John Wright, Bristol.
10. Goffman, I. (1961). *Asylums*. Doubleday, New York; Penguin, Harmondsworth, 1968.
11. Clark, D.H. (1965). The therapeutic community—concept, practice and future. *The British Journal of Psychiatry*, **131**, 553–64.
12. Campling, P., Davies, S., and Farquharson, G. (2004). *From toxic institutions to therapeutic environments*. Gaskell, London.
13. Hardcastle, M., Kennard, D., Grandison, S., *et al.* (2007). *Experiences of mental health in-patient care: narratives from service users, carers and professionals*. Routledge, London.
14. Ciompi, L. and Hoffman, H. (2004). Soteria Berne: an innovative milieu therapeutic approach to acute schizophrenia based on the concept of affect-logic. *World Psychiatry*, **3**, 140–6.
15. Stanton, A. and Schwartz, H. (1954). *The mental hospital*. Basic Books, New York.
16. Menzies, I. (1960). A case-study in the functioning of social systems as a defence against anxiety. *Human Relations*, **13**, 95–121.
17. Main, T. (1957). The ailment. *British Journal of Medical Psychology*, **30**, 129–45.
18. Haigh, R. (1999). The quintessence of a therapeutic environment. In *Therapeutic communities: past, present and future* (eds. P. Campling and R. Haigh), Chap. 20. Jessica Kingsley, London.
19. Jones, M. (1968). *Social psychiatry in practice*, pp. 105–12. Penguin, Harmondsworth.
20. Main, T. (1983). The concept of the therapeutic community: variations and vicissitudes. In *The evolution of group analysis* (ed. M. Pines), pp. 197–217. Routledge and Kegan Paul, London.
21. Roberts, J. (1998). Questions of training. In *An introduction to therapeutic communities* (ed. D. Kennard). Jessica Kingsley, London.
22. Rawlings, B. (2005). The temporary therapeutic community—a qualitative evaluation of an ATC training weekend. *Therapeutic Communities*, **26**, 6–18.
23. Mahoney, N. (1979). My stay at the Henderson Therapeutic Community. In *Therapeutic communities: reflections and progress* (eds. R.D. Hinshelwood and N. Manning), pp. 76–87. Routledge and Kegan Paul, London.
24. Wexler, H. (1997). Therapeutic communities in American prisons. In *Therapeutic communities for offenders* (eds. E. Cullen, L. Jones, and R. Woodward), pp. 161–79. Wiley, Chichester.
25. Lees, J., Manning, N., and Rawling, B. (1999). *Therapeutic community effectiveness. A systematic international review of therapeutic community treatment for people with personality disorders and mentally disordered offenders*. CRD Report 17, NHS Centre for Reviews and Dissemination, University of York, York.
26. Warren, F., Preedy-Fayers, K., McGauley, G., *et al.* (2003). *Review of treatments for severe personality disorder*. Home Office Online Report 30/03, Home Office, London.
27. Rawlings, B. (1999). Therapeutic communities in prisons: a research review. *Therapeutic Communities*, **20**, 177–93.
28. Yates, R. and Wilson, J. (2001). The modern therapeutic community: dual diagnosis and the problem of change. In *Therapeutic communities for the treatment of drug users* (eds. B. Rawlings and R. Yates). Jessica Kingsley Publishers, London and Philadelphia.
29. Leff, J. and Trieman, N. (1997). Providing a comprehensive community psychiatric service. In *Care in the community: illusion or reality* (ed. J. Leff), pp. 189–201. Wiley, Chichester.

30. Sacks, S. (2000). Co-occurring mental and substance use disorders: promising approaches and research issues. *Substance Use and Misuse*, **35**, 2061–93.
31. Rose, M. (1997). *Transforming hate to love: an outcome study of the Peper Harow treatment process for adolescents*. Routledge, London.
32. Wright, J.C. and Richardson, P. (2003). The challenge of research. In *Therapeutic communities for children and young people* (eds. A. Ward, K. Kasinski, J. Pooley, and A. Worthington), pp.244–53. Jessica Kingsley, London.
33. Cox, C. (1995). The case for village communities for people with learning disabilities. *British Journal of Nursing*, **4**, 1130–4.

6.4

Treatment by other professions

Contents

6.4.1 Rehabilitation techniques

W. Rössler

The goal of psychiatric rehabilitation is to help disabled individuals to establish the emotional, social, and intellectual skills needed **to live, learn, and work in the community** with the least amount of professional support.(1)

Rehabilitation practice has changed the perception of mental illness. Enabling disabled people to live a normal life in the community causes a shift away from a focus on an illness model towards **a model of functional disability**. As such, other outcome measures aside from clinical conditions become relevant. Social role functioning including social relationship, work, and leisure as well as quality of life and family burden are of major interest for the people affected living in the community.(25)

The relevance of psychosocial and environmental problems is reflected in the DSM-IV and ICD-10. Axis IV of DSM-IV and codes Z55–Z65 and Z73 of ICD-10 are assigned for reporting psychosocial and environmental problems that may affect the diagnosis, treatment, and prognosis of mental disorders.

The International Classification of Functioning, Disability and Health

Long-term consequences of major mental disorders might be described using different dimensions. A useful tool was provided by the International Classification of Impairment, Disability and Handicaps (ICIDH), first published by the World Health Organization in 1980. The ICIDH has been recently revised. The revised '**International Classification of Functioning, Disability and Health'** (ICF) includes a change from negative descriptions of impairments, disabilities and handicaps to neutral descriptions of body structure and function, activities and participation. A further change has been the inclusion of a section on environmental factors as part of the classification. This is in recognition of the importance of the role of environmental factors in either facilitating functioning or creating barriers for people with disabilities. Environmental factors interact with a given health condition to create a disability or restore functioning, depending on whether the environmental factor is a facilitator or a barrier.

ICF is a useful tool to comprehend chronically mentally ill in all their dimensions including impairments at the structural or functional level of the body, at the person level concerning activity limitations and at the societal level with respect to restrictions of participation. Each level encompasses a theoretical foundation on which a respective rehabilitative intervention can be formulated.

Target population

During the course of psychiatric reforms the predominant objective of psychiatric rehabilitation was to resettle patients from large custodial institutions to community settings. Today all patients suffering from severe mental illness require rehabilitation. The core group is drawn from patients with the following:

- persistent psychopathology
- marked instability characterized by frequent relapse
- social maladaptation.(28)

There are other definitions currently used to characterize the chronically mentally ill. But they all share some common elements, that is a diagnosis of mental illness, prolonged duration, and **role incapacity**.

Although the majority of the chronically mentally ill have the diagnosis of schizophrenic disorders, other patient groups with psychotic and non-psychotic disorders are targeted by psychiatric rehabilitation. Up to 50 per cent of people with severe mental illness carry **dual diagnoses** especially in combination with substance abuse.

The role of the psychiatrist in rehabilitation

Psychiatric rehabilitation is by its very nature, **multidisciplinary;** because of the many different competencies required. Monitoring medication is a key task of the psychiatrist.

Pharmacotherapy in psychiatric rehabilitation needs some special consideration. Symptom control does not necessarily have the highest priority as some side effects of pharmacological treatment can weaken a person's ability to perform his social roles, and impair vocational rehabilitation. Many patients living in the community **want to take responsibility for their medication** themselves. This also includes the varying of medication without consultation within certain limits.

The starting point for an adequate understanding of rehabilitation is that it is concerned with the individual person in the context of his or her specific environment. Psychiatric rehabilitation is regularly **carried out under real life conditions**. Thus, rehabilitation practitioners have to take into consideration the realistic life circumstances that the affected persons are likely to encounter in their day-to-day living.

A necessary second step is helping disabled persons to **identify their personal goals**. Motivational interviews provide a sophisticated approach to identify the individuals' personal costs and benefits associated with the needs listed.[8] This makes it also necessary to assess the individuals' readiness for change.[15] Functional assessment and individual goal setting are prerequisites of a differentiated rehabilitation intervention plan and should be repeated in different stages of the rehabilitation process.

The rehabilitative planning process **focuses on the patient's strengths**. Irrespective of the degree of psychopathology of a given patient, the rehabilitation practitioner must **work with the 'well part of the ego'** as 'there is always an intact portion of the ego to which treatment and rehabilitation efforts can be directed' (Lamb 1984[2]). This leads to a closely related concept: the aim of restoring hope to people who suffered major setbacks in self-esteem because of their illness. As Bachrach[2] states 'it is the kind of hope that comes with learning to accept the fact of one's illness and one's limitations and, proceeding from there'.

Psychiatric rehabilitation **concentrates on peoples' rights as a respected partner** and **endorses their involvement and self-determination** concerning all aspects of the treatment and rehabilitation process. These rehabilitation values are also incorporated in the **concept of recovery.**[9] Within the concept of recovery, the **therapeutic alliance** plays a crucial role in engaging the patient in his or her own care planning. It is essential that the patient can rely on his or her therapist's understanding and trust as most of the chronically mentally ill and disabled persons lose close, intimate, and stable relationships in the course of the disease. Recent research has suggested that social support is associated with recovery from chronic diseases, greater life satisfaction, and enhanced ability to cope with life stressors.[24] Therefore, psychiatric rehabilitation is also an exercise in network building.

Current approaches

Psychiatric rehabilitation aims at **changing the natural course** of the disease. Yet, there is no consensus among rehabilitation researchers on what rehabilitation actually does accomplish. Some understand rehabilitation as an approach to help disabled people to compensate for impairments and to function optimally with the deficiencies they have. Other researchers assume that rehabilitation helps the patient to recover from the disorder itself, while the contributing factors to the healing process are not clear.

The overall philosophy of psychiatric rehabilitation comprises two intervention strategies. The first strategy is individual-centered and aims at **developing the patient's skill** to interact with a stressful environment. The second strategy is ecological and is directed towards **developing environmental** resources to reduce potential stressors. Most disabled people need a combination of both approaches.

As a general rule people with psychiatric disabilities tend to have the **same life aspirations** as people without disabilities in their society or culture. They want to be respected as autonomous individuals and lead a life as normal as possible. As such they mostly desire (1) their own housing, (2) an adequate education and a meaningful work career, (3) satisfying social and intimate relationships, and (4) participation in community life with full rights.

Housing

The objective of psychiatric reforms since the mid-50s of the 20th century has been to resettle chronically mentally ill persons from large custodial institutions to community settings. Providing sheltered housing in the community for the long-term patients of the old asylums was one of the first steps in the process of deinstitutionalization. Most long-stay patients can successfully **leave psychiatric hospitals and live in community settings**.

Ideally, a **residential continuum** (RC) with different housing options should be provided. RC ranges from round-the-clock staffed sheltered homes to more independent and less staffed sheltered apartments, which eventually allow individuals moving to independent housing in the community. Critics of RC contended that (1) up to date RC is rarely available in communities, (2) that RC does not meet the varying and fluctuating needs of persons with serious mental illnesses, and (3) that RC does not account for individuals' preferences and choices. **Supported housing**, i.e. independent housing coupled with the provision of support services emerged in the 1980s as an alternative to RC. Supported housing offers flexible and individualized services depending on the individual's demands. In the meantime, rehabilitation research could demonstrate that supported housing is a realistic goal for the majority of people with psychiatric disabilities.[23] Once in supported housing, the majority stay in housing and are less likely to become hospitalized. Other outcomes do not yield consistent results.

Work

The beneficial effects of work on mental health have been known for centuries.[11] Therefore, **vocational rehabilitation** has been

a core element of psychiatric rehabilitation since its beginning. Vocational rehabilitation is based on the assumption that work not only improves activity, social contacts, etc., but may also promote gains in related areas such as self-esteem and quality of life, as work and employment are a step away from dependency and a step closer to integration into society. **Enhanced self-esteem** in turn improves adherence to rehabilitation of individuals with impaired insight.

Vocational rehabilitation originated in psychiatric institutions where the lack of activity and stimulation led to apathy and withdrawal of their inpatients. Long before the introduction of medication, occupational and work therapy contributed to sustainable improvements in long-stay inpatients. Today occupational and work therapy are not any longer hospital-based but represent the starting point for a wide variety of rehabilitative techniques teaching vocational skills.

Vocational rehabilitation programs in the community provide a series of **graded steps to promote job entry or re-entry**. For less disabled persons, brief and focused techniques are used to teach how they can find a job, fill out applications, and conduct employment interviews. In transitional employment, a temporary work environment is provided to teach vocational skills, which should enable the affected person to move on to competitive employment. But all too often, the gap between transitional and competitive employment is so wide that the mentally disabled individuals remain in a temporary work environment. Sheltered workshops providing pre-vocational training also quite often prove a dead end for the disabled persons.

One consequence of the difficulties in integrating mentally disabled individuals into the common labour market has been the steady growth of cooperatives, which operate commercially with disabled and non-disabled staff working together on equal terms and sharing in management. The mental health professionals work in the background providing support and expertise.

Today, the most promising vocational rehabilitation model is **supported employment** (SE). In SE, disabled persons are placed in competitive employment according to their choices as soon as possible, and receive all support needed to maintain their position.[4] The support provided is continued indefinitely. Participation in SE programs is followed by an increase in the ability to find and keep employment.[7] Links were also found between job tenure and non-vocational outcomes, such as improved self-esteem, social integration, relationships, and control of substance abuse.[4,29] It was also demonstrated that those who had found long-term employment through SE had improved cognition, quality of life, and better symptom control.[17]

Although findings regarding SE are encouraging, some critical issues remain to be answered. Many individuals in SE obtain unskilled part-time jobs. Since most studies only evaluated short (12–18 months) follow-up periods, the long-term impact remains unclear. Currently, we do not know which individuals benefit from supported SE and which do not.[20] After all, we have to realise that the integration into the labour market does by no means only depend on the ability of the persons affected to fulfill a work role and on the provision of sophisticated vocational training and support techniques but also on the **willingness of society to integrate** its most disabled members.

Building relationships

In recent years, social skills training packaged in the form of modules with different topics has become very popular and has been widely promulgated. The modules focus on medication management, symptom management, substance abuse management, basic conversational skills, interpersonal problem solving, friendship and intimacy, recreation and leisure, workplace fundamentals, community (re-)entry, and family involvement. Each module is composed of skill areas. The skills areas are taught in exercises with demonstration videos, role-play and problem solving exercises, and *in-vivo* and homework assignments.[14]

The results of several control studies suggest that disabled individuals **can be taught a wide range of social skills**. Social and community functioning improve when the trained skills are relevant for the patient's daily life, and the environment perceives and reinforces the changed behaviour. Unlike medication effects, benefits from skills training occur more slowly. Furthermore, long-term training has to be provided for positive effects.[3] Overall, social skills training have been shown to be effective in the acquisition and maintenance of skills and their transfer to community life.[13]

Keeping relationships

As a consequence of deinstitutionalization the **burden of care** has increasingly fallen on the relatives of the mentally ill. Informal caregiving significantly contributes to health care and rehabilitation.[31] Fifty to ninety per cent of disabled persons live with their relatives following acute psychiatric treatment. This is a task many families do not choose voluntarily. Caregiving imposes a significant burden on families. Those providing informal care face considerable adverse health effects, including higher levels of stress and depression, and lower levels of subjective well being, physical health and self-efficacy. Additionally, not all families are equally capable of giving full support to their disabled member and are not willing to replace an insufficient health care system. Caregivers regularly experience higher levels of burden when they have poor coping resources and reduced social support. But **families also represent support systems**, which provide natural settings for context-dependent learning important for recovery of functioning. As such, there has been a growing interest in helping affected families since the beginning of care reforms.

One area of interest deals with the expectations of relatives concerning the provision of care. Relatives quite **often feel ignored**, not taken seriously, and also feel insufficiently informed by health professionals. They also may feel that their contribution to care is not appreciated or that they will be blamed for any patient problems. It certainly is no surprise that there is a lot of frustration and resentment among relatives considering the physical, financial and emotional family burden.

Family intervention programs have produced promising results. Family intervention is effective in lowering relapse rate, and also in improving outcome e.g., psychosocial functioning. Possibly, family intervention can reduce family burden. Furthermore, the treatment gains are fairly stable.[21] But we also have to appreciate, that it is not clear what the effective components of the different models are. Additionally, family interventions differ in frequency and length of treatment. There are also no criteria for the minimum amount of treatment necessary.

Finally, we have to be aware that most family interventions were developed in the context of western societies during deinstitutionalization. Family caregiving might be quite different in a **different cultural context**. This refers to other cultures in total as well as to minority groups in western societies.[31]

Participation in community life with full rights

Practitioners often are confronted with the **deleterious effects of stigma and discrimination** in the lives of people with serious mental illnesses. Numerous studies have examined stigmatizing attitudes toward people with mental illness.[12] In recent years, the scientific interest in the perspective of the labelled individual has increased too. There is extensive empirical evidence of the negative consequences of labelling and perceived stigmatization. These include demoralization, low quality of life, unemployment, and reduced social networks.[10,19] Once assigned the label 'mental illness' and having become aware of the related negative stereotypes, the affected individuals expect to be rejected, devaluated or discriminated against. This vicious cycle decreases the chance of recovery and normal life.

On the other hand, well-integrated people with mental illness exhibit better outcomes regarding psychopathology and quality of life. The importance of social integration is underlined even more when considering the subjective availability of support: perceived social support predicts outcome in terms of recovery from acute episodes of mental illness, community integration, and quality of life.[27,30]

On the basis of comprehensive research in this area during the last decade, several strategies have been developed to fight the stigma and discrimination suffered by those who have mental illnesses.[30] Different research centres developed interventions directed to specific target groups relevant for de-stigmatization, e.g., students[18] or police officers.[22] Persons in contact with mentally ill individuals quite often have a more positive attitude. Contact with the mentally ill persons also reduces social distance,[12] which is a strong argument in favour of community psychiatry. Other initiatives have targeted stigma by means of more comprehensive programs. The World Psychiatric Association launched one of the internationally best-known programs in 1996 (www.openthedoors.com). All these initiatives make clear that efforts in re-integrating persons with serious mental illness into community life must be accompanied by measures on the societal level.

The core elements of modern psychiatric rehabilitation are summarized in Box 6.4.1.1

Developing environmental resources

Effective psychiatric rehabilitation requires **individualized and specialized treatment**, which has to be embedded in a comprehensive and coordinated system of rehabilitative services. But even when a variety of services are available, they are poorly linked in many cases, and costly duplication may occur.

While developing community support systems it became obvious that there is a **need to coordinate and integrate the services** provided as each involved professional concentrates on different aspects of the same patient. Therefore, as a key coordinating and integrating mechanism, the concept of **case management** (CM) originated. CM focuses on all aspects of the physical and social environment. The core elements of CM are the assessment of patients' needs, the development of comprehensive service plans for the patients, and arrangement of service delivery.[26]

Over the past two decades, a variety of different models of CM have been developed which exceed the original idea that CM mainly intends to link the patient to needed services and to coordinate those services. Today most clinical case managers also provide direct services in the patient's natural environment. This model is called **Intensive Case Management** (ICM). ICM on its part is difficult to distinguish from **Assertive Community Treatment** (ACT).

Stein and Test have developed the basic compounds of ACT in the 1970's.[33] The original program was designed as a community based alternative to hospital treatment for persons with severe mental illnesses. A comprehensive range of treatment, rehabilitation, and support services in the community is provided through a multidisciplinary team. ACT is characterized by an assertive outreach approach i.e., interventions are mainly provided in the natural environment of the disabled individuals.[32]

Research on CM and ACT yielded 'mixed' results.[6] While the traditional office-based CM approach obviously is less successful, the ACT model was found to be more beneficial when compared with standard care.[16] ACT can reduce time in hospital,[20] but has moderate or only little effects on improving symptomatology and social functioning. The differing features of the respective services might explain the international variation. Six regularly occurring features of successful services were identified: smaller case loads, regularly visits at home, a high percentage of contacts at home, responsibility for health and social care, multidisciplinary teams, and a psychiatrist integrated in the team.[5]

Box 6.4.1.1 Core elements of psychiatric rehabilitation

- multidisciplinary
- carried out under real life conditions
- identify patient's personal goals
- focus on the patient's strengths
- work with the 'well part of the ego,
- concentrate on people's rights as a respected partner
- endorse their involvement and self-determination
- build a therapeutic alliance

References

1. Anthony, W. (1979). *The principles of psychiatric rehabilitation*. University Park Press, Baltimore, MD.
2. Bachrach, L.L. (2000). Psychosocial rehabilitation and psychiatry in the treatment of schizophrenia – what are the boundaries? *Acta Psychiatr Scand Suppl* **407**, 6–10.
3. Bellack, A.S. (2004). Skills training for people with severe mental illness. *Psychiatr Rehabil J*, **27**(4): 375–91.
4. Bond, G.R. (2004). Supported employment: evidence for an evidence-based practice. *Psychiatr Rehabil J*, **27**(4): 345–59.
5. Burns, T, Catty, J, Wright, C. (2006). De-constructing home-based care for mental illness: can one identify the effective ingredients? *Acta Psychiatr Scand Suppl*, **429,** 33–5.
6. Burns, T, Fioritti, A, Holloway, F. *et al*. (2001). Case management and assertive community treatment in Europe. *Psychiatr Serv*, **52**(5), 631–6.
7. Cook, J.A., Leff, H.S., Blyler, C.R. *et al*. (2005). Results of a multisite randomized trial of supported employment interventions for individuals with severe mental illness. *Arch Gen Psychiatry,* **62**(5), 505–12.
8. Corrigan, P.W., McCracken, S.G., Holmes, E.P. (2001). Motivational interviews as goal assessment for persons with psychiatric disability. *Community Ment Health J*, **37**(2), 113–22.

9. Farkas, M., Gagne, C., Anthony, W. *et al.* (2005). Implementing recovery oriented evidence based programs: identifying the critical dimensions. *Community Ment Health J,* **41**(2), 141–58.
10. Graf, J., Lauber, C., Nordt, C. *et al.* (2004). Perceived stigmatization of mentally ill people and its consequences for the quality of life in a Swiss population. *J Nerv Ment Dis*, **192**(8), 542–7.
11. Harding, C., Strauss, J., Hafez, H. *et al.* (1987). Work and mental illness. I. Toward an integration of the rehabilitation process. *Journal of Nervous and Mental Disease*, **175**, 317–26.
12. Lauber, C., Nordt, C., Falcato, L. *et al.* (2004). Factors influencing social distance toward people with mental illness. *Community Ment Health J*, **40**(3), 265–74.
13. Liberman, R.P., Glynn, S., Blair, K.E. *et al.* (2002). In vivo amplified skills training: promoting generalization of independent living skills for clients with schizophrenia. *Psychiatry,* **65**(2), 137–55.
14. Liberman, R.P., Kopelowicz, A. (2002). Teaching persons with severe mental disabilities to be their own case managers. *Psychiatr Serv* **53**(11), 1377–9.
15. Liberman, R.P., Wallace, C.J., Hassell, J. (2004). Rehab rounds: Predicting readiness and responsiveness to skills training: the Micro-Module Learning Test. *Psychiatr Serv*, **55**(7), 764–6.
16. Marshall, M. (1996). Case management: a dubious practice. BMJ, **312**(7030), 523–4.
17. McGurk, SR., Mueser, K.T. (2003). Cognitive functioning and employment in severe mental illness. *J Nerv Ment Dis*, **191**(12), 789–98.
18. Meise, U., Sulzenbacher, H., Kemmler, G. *et al.* (2000). ["…not dangerous, but nevertheless frightening". A program against stigmatization of schizophrenia in schools]. *Psychiatr Prax*, **27**(7), 340–6.
19. Mueller, B., Nordt, C., Lauber, C. *et al.* (2006). Social support modifies perceived stigmatization in the first years of mental illness: a longitudinal approach. *Soc Sci Med* 2006, **62**(1), 39–49.
20. Mueser, K.T., Bond, G.R., Drake, R.E. *et al.* (1998). Models of community care for severe mental illness: a review of research on case management. *Schizophrenia Bulletin*, **24**, 37–74.
21. Pilling, S., Bebbington, P., Kuipers, E. *et al.* (2002). Psychological treatments in schizophrenia: I. Meta-analysis of family intervention and cognitive behaviour therapy. *Psychol Med*, **32**(5), 763–82.
22. Pinfold, V., Huxley, P., Thornicroft, G. *et al.* (2003). Reducing psychiatric stigma and discrimination–evaluating an educational intervention with the police force in England. *Soc Psychiatry Psychiatr Epidemiol*, **38**(6), 337–44.
23. Rog, DJ. (2004). The evidence on supported housing. *Psychiatr Rehabil J,* **27**(4), 334–44.
24. Rogers, E.S., Anthony, W., Lyass, A. (2004). The nature and dimensions of social support among individuals with severe mental illnesses. *Community Ment Health J*, **40**(5), 437–50.
25. Rössler, W. (2006). Psychiatric rehabilitation today: an overview. World Psychiatry **5**(3), 151–7.
26. Rössler, W., Fätkenheuer, B., Löffler, W. *et al.* (1992). Does case management reduce the rehospitalization rate? *Acta Psychiatrica Scandinavica*, **86**, 445–9.
27. Rössler, W., Salize, H.J., Cucchiaro, G. *et al.* (1999). Does the place of treatment influence the quality of life of schizophrenics? *Acta Psychiatr Scand*, **100**(2), 142–8.
28. Royal College of Psychiatrists (1996). *Psychiatric rehabilitation* (revised edition). Gaskell, London.
29. Ruesch, P., Graf, J., Meyer, P.C. *et al.* 2004). Occupation, social support and quality of life in persons with schizophrenic or affective disorders. *Soc Psychiatry Psychiatr Epidemiol*, **39**(9), 686–94.
30. Rusch, N., Angermeyer, M.C., Corrigan, P.W. (2005). Mental illness stigma: concepts, consequences, and initiatives to reduce stigma. *Eur Psychiatry,* **20**(8), 529–39.
31. Schulze, B., Rössler, W. (2005). Caregiver burden in mental illness: review of measurement, findings and interventions in 2004–2005. *Current Opinion in Psychiatry*, **18**, 684–91.
32. Scott, J.E. and Dixon, L.B. (1995). Assertive community treatment and case management for schizophrenia. *Schizophrenia Bulletin*, **21**, 657–68.
33. Stein, L.I. and Test, M.A. (1980). Alternative to mental hospital treatment. I. Conceptual model, treatment program, and clinical evaluation. *Archives of General Psychiatry*, **37**, 392–7.

6.4.2 Psychiatric nursing techniques

Kevin Gournay

Background

Psychiatric nursing as an entity has really only evolved since the Second World War. Psychiatric nurses (now often referred to as mental health nurses in the United Kingdom and Australasia) can now be found in most countries of the developed world, although in the developing world, psychiatric nursing is still not defined as a specific discipline. In many countries, psychiatric hospitals are still staffed by untrained 'Attendants' who may have some supervision from general trained nurses. Nevertheless, a number of initiatives, notably those of the Geneva Initiative in Psychiatry[(1)] in Eastern Europe and the former Soviet Union and the World Health Organization in African countries, have provided specific training in psychiatric nursing techniques.

The development of psychiatric nursing across the world needs to be seen in the context of changing and evolving patterns of mental health care. De-institutionalization, with the attendant setting up of community mental health teams, has prompted a range of innovations in psychiatric nursing and the psychiatric nurse of today, who in the United States and Europe is likely to be a university graduate, is a very different person to that of the nurse working in the post-Second World War asylums of 40 years ago.

In this chapter, we examine the development of psychiatric nursing in some detail and particularly emphasize the role of psychiatric nurses working in the community. Community psychiatric nursing first developed in the United Kingdom nearly 50 years ago and this model has been followed in countries such as Australia and New Zealand. However, this community role has not developed to any great extent in the United States, where the main presence of psychiatric nursing remains in hospital-based care. Furthermore, in the United Kingdom and Australasia, the development of community initiatives has seen the role of the psychiatric nurse blurring with that of other mental health professionals. Chapters such as this cannot really do justice to the whole range of techniques used by psychiatric nurses; neither can it examine in any detail the differences between psychiatric nursing practices across the world. However, a description of psychiatric nursing in six important areas will provide the reader with an appreciation of the range and diversity of psychiatric nursing skills:

- Inpatient care
- Psychosocial interventions in the community
- Prescribing and medication management

- Cognitive behaviour therapy
- Primary care
- Psychiatric nursing in the developing world.

Psychiatric nursing in inpatient settings

In the past three decades the population in psychiatric hospitals across the developed world has fallen dramatically in England from 160 000 to 30 000 beds over a period of 25 years and the duration of inpatient care in the United Kingdom in 2007 is approximately 36 days. However, today's inpatients are a population with much greater levels of illness than was previously the case; they tend to be more treatment-resistant, have complex problems, and display high levels of substance abuse and violence.[2] As a corollary of this, a greater proportion of patients are now detained under mental health legislation. Inpatient facilities consist of acute psychiatric units, local secure units, and high secure psychiatric hospitals for those patients who pose high levels of danger to themselves and others. In the United Kingdom, four high secure hospitals contain approximately 1600 patients. It should also be noted that, due to the large numbers of people with psychiatric problems in prisons, there are now several hundred psychiatric nurses employed in prison settings to carry out a range of assessment and treatment procedures. In addition, the NHS also has a number of 'in reach' schemes, which include sending NHS staff into prisons on a sessional basis.

Given that community care in the western World is now the norm, inpatient care is now seen as a short-term measure with the dual purposes of stabilizing the patient's condition and keeping the patient safe. Psychiatric nurses have a role to play in the overall assessment of the patient and, given that the nurse is—literally—with the patient 24 h a day, the observation of the patient's mental state and behaviour is of considerable importance. Unfortunately, this is an area where, outside the United States, a number of problems exist and suicide rates by inpatients are unacceptably high.[3] In the United States, inpatient wards tend to be much more secure than wards in countries such as the United Kingdom and Australia and, therefore, the incidence of inpatient suicide is much lower. The *UK National Confidential Inquiry into suicides and homicides*[3] demonstrates that nearly 200 suicides by inpatients occur every year, with hanging on the ward itself being still prevalent at unacceptably high levels. Recently the National Institute for Health and Clinical Excellence (NICE)[4] has published guidelines, which include the observation of patients at risk. This guidance sets out very careful protocols for the observation of patients at risk and includes recommendations regarding the prevention of absconding. In the United Kingdom and Australia, open-door policies still operate in acute psychiatric units and it is being increasingly recognized that balancing the rights of the patients against safety is a difficult issue. Nurses also have a major role to play in providing patients and their families with information about condition and treatment. We also know that there are interventions that can be applied by nurses, which would lead to improved outcomes. For example, Drury *et al.*[5] showed that a cognitive behavioural therapy package improved longer-term outcomes. Similarly, Kemp *et al.*[6] showed that motivational interviewing and psychoeducation methods produced clear, clinical, and economic benefits in patients who have compliance problems with medication.

With regard to the containment of violent behaviour, which is now so common in inpatient settings, nurses in the United Kingdom have been assisted by very comprehensive evidence-based guidance from the National Institute for Health and Clinical Excellence (NICE),[4] which sets out clear guidance on the use of de-escalation techniques and control and restraint, as well as providing a comprehensive algorithm for the use of rapid tranquillization. In respect of rapid tranquillization, nurses are now provided with the necessary skills to observe and monitor patients following rapid tranquillization, including the use of pulse oximetry and blood pressure. Whilst nurses in Australasia use the same methods of managing violent behaviour as nurses in the United Kingdom, psychiatric nurses in most European countries and in the United States use various forms of mechanical restraint and a very wide range of devices, including belts, straps, nets, and jackets. Whilst it needs to be recognized that there are a range of social and cultural influences that determine how violence in mentally ill people is managed, it is important to note that the evidence base for all forms of violence management, including rapid tranquillization, is very poor and a Cochrane review found that there is no evidence base for the use of seclusion and restraint.[7]

Psychosocial interventions in the community

In order to appreciate the current practice of psychiatric nurses working in the community, it is important to say something about the historical context. Until the early 1980s, community psychiatric nurses (CPNs) in the United Kingdom were generally based in large, Victorian psychiatric hospitals and worked mostly within a consultant psychiatrist team responsible for the follow-up of patients after discharge from hospital. Their main responsibilities were the administration of medication and the provision of general, supportive care, mostly to people with schizophrenia, the elderly with functional and organic illnesses, and to people with other serious and enduring mental illnesses. Initial research on the effectiveness of community psychiatric nurses produced very positive results. In a randomized trial conducted by Paykel *et al.*[8] CPNs were compared with psychiatric registrars in the provision of aftercare for patients who had suffered an acute episode requiring hospitalization. In general terms, this study showed that there was an equivalent outcome on clinical, social, and economic measures. Some 20 years ago, CPNs in the United Kingdom began to diversify their practice and separated themselves from consultant psychiatrists, attaching themselves to primary care settings and taking referrals directly from GPs. By 1990, a national survey showed that 40 per cent of CPNs worked in primary care.[9] The vast majority of this work involved treating people with depression, anxiety, and adjustment disorder, using counselling-based approaches. Whilst this work by CPNs became very popular with GPs and mental health professionals in general, research into the effectiveness of their work demonstrated that they were largely ineffective. Gournay and Brooking[10] carried out a randomized controlled trial involving 11 CPNs, working in six primary care settings in North London. In this study, 177 patients were randomized to either routine continuing care from their GP or to CPN intervention. The majority of patients had adjustment disorders and various states of general depression and anxiety. Patients, in both the CPN and

continuing GP care groups, showed significant improvement on a range of measures, clinical status, and social functioning but, at post-treatment and follow-up, there was no difference in outcomes demonstrated. Patients allocated to CPNs showed high levels of dropout (50 per cent) and patient satisfaction rating did not correlate with outcome measures. An economic analysis[11] showed that, per unit of health gain, CPN intervention was very expensive compared with interventions for people with schizophrenia. The Paykel and Gournay and Brooking studies still represent the only research evidence regarding the efficacy of CPNs working with common mental disorders.

During the early 1990s a National Review of Psychiatric Nursing in the United Kingdom led to CPNs refocusing their efforts on the seriously mentally ill and this trend has been followed in Australasia. In the last decade there has been a wide range of psychiatric nursing developments in respect of psychosocial interventions. The initial impetus for this development came from the Thorn Programme, this initiative taking its name from the Sir Jules Thorn Trust, a charitable foundation that provided the funds to inaugurate the first 3 years of the training programme for nurses, commencing in 1992. The initiative was originally led by Dr Jim Birley who, with a group of colleagues from other professions, became impressed by the work of nurses working in cancer care. Birley's initial aim was to train a substantial number of nurses specifically dedicated to the care of people with schizophrenia and their relatives. Indeed, Birley, who was one of the pioneers of Social Psychiatry in the United Kingdom, noted that the families of people with schizophrenia were often in great need of intervention. Previous work in Manchester[12] had confirmed that nurses could be trained in family intervention skills which in turn led to positive outcomes for the patient and family. This training in family work formed the basis of what has now become a more general initiative to train nurses in various evidence-based psychosocial interventions for schizophrenia. The Thorn Initiative has now become the national model of training in psychosocial interventions in the United Kingdom and similar programmes to Thorn have been set-up in Australasia and some European countries. The psychosocial interventions used by nurses are as follows:

- Assertive community treatment
- Family interventions for schizophrenia
- Cognitive behavioural techniques for managing hallucinations and delusions
- Approaches with dual diagnosis
- Medication management.

In addition to psychosocial interventions, training programmes for psychiatric nurses working in the community, now also include approaches to improve the physical health of people with serious mental health problems and nurses are now taking a more active lead in ensuring that this very vulnerable population obtains appropriate medical services including physical screening and health promotion activities. At the time of writing, there are also, in several parts of the United Kingdom, specific training programmes for nurses aimed at helping patients with chronic mental illness to deal with obesity, lack of exercise, and smoking.

On a cautionary note, there are now several studies[13] which demonstrate that more intensive case management may not be effective, the possible reason being that one needs to provide nurses with suitable levels of training in community approaches. In the UK700 study[13] mentioned above the nurses involved only received a few hours training in case management approaches, whist nurses undertaking basic psychosocial interventions training such as a Thorn diploma will receive 250 h of classroom instruction in addition to supervised practice. Whilst there have been considerable numbers of nurses trained in the above-mentioned evidence-based psychosocial interventions, unfortunately there are still many nurses working in the community without such training. Whilst their general psychiatric nursing skills will be reasonably sound, their impact on patient care will be somewhat limited.

In the early part of the twenty-first century, psychiatric nurses working in the United Kingdom, Europe, and Australasia are increasingly working in specialist community teams, for example, assertive outreach services, early intervention teams and crisis intervention, and home treatment teams. Whilst these approaches are commonly used across the United States, such teams are likely to be staffed by case managers who have backgrounds in social work and social care and psychiatric nurses are unlikely to be employed in large numbers. In the United States, psychiatric nurses are often specifically employed to run medication clinics, probably for reasons of cost, whilst in the United Kingdom, approximately 50 per cent of community mental health teams carrying out a very wide range of psychosocial functions are likely to be CPNs.

Prescribing and medication management

The work of Kemp *et al.*[6] who showed that motivational interviewing and psychoeducation methods produced good outcomes for patients who were non-compliant with their medication, led to the development of medication management training for nurses in the United Kingdom. Gray *et al.*[14] using a cluster randomized controlled trial, where 60 CPNs were randomly assigned to medication management training or carrying on with their usual treatment as usual, demonstrated that the nurses who had received medication management training produced very clear benefits in patients with schizophrenia. The study demonstrated a significantly greater reduction in patients' overall psychopathology for the trained group, compared with treatment as usual. At the end of the 6-month study period, the improvement in positive and negative symptoms for the trained nurses over the control was statistically and clinically significant. This training is now used by, literally, thousands of nurses in the United Kingdom, Australasia, and some non-English speaking countries in Europe. Whilst this training, which comprised a number of components, including improving the pharmacological knowledge of the nurses, the use of side effect monitoring and motivational interviewing for non-compliant patients, a European multi-centre trial, which tested adherence therapy as a treatment package over and above routine clinical care and delivered mostly by psychologists and psychiatrists, showed that the treatment package was no more effective than health education in improving quality of care.[15] Both studies raised a number of questions concerning the very complex issue of treatment compliance with medication and the specific difficulties associated with measurement of compliance itself and of patient insight.

Arguably, the most important recent development in psychiatric nursing has been the advent of nurse prescribing. This began more than a decade ago in the United States, where nurses have

prescriptive authority in virtually all states. The situation in the United States is, however, complex with a considerable variation in the level of prescriptive authority across the United States from complete independence to being able to prescribe under a physician protocol. In turn, the educational requirement for nurse prescribers also varies considerably. However, most states have fairly comprehensive regulations concerning not only course content, but also hours of instruction and supervision. The Website of the American Psychiatric Nurses,[16] provided at the end of this chapter, provides very detailed state-by-state information. In 2006, United Kingdom legislation was passed that means nurses may prescribe almost independently,[17] although—as in the United States—there is a variation across the country in terms of training and practice. Across nursing more generally, the law in the United Kingdom means that provided it is within their area of specialist work, nurses may independently prescribe any drug (including controlled substances such as opiates). By contrast to the different legislative frameworks that exist in the United States, the legislative framework for the United Kingdom is unitary. However, interpretation of that framework in the United Kingdom seems to vary between NHS services. There are now similar nurse prescribing initiatives in Australasia, where because of the nature of rural and remote populations, the development of nurse prescribing seems very logical.

At present, there are no randomized controlled trial data to compare nurse prescribing with more conventional doctor prescribing, neither is there any substantial data on patient efficacy. The advantages of nurse prescribing have been clearly set out in a *Maudsley discussion paper* by Gournay and Gray[18] and these include the delegation of routine prescribing tasks to nurses, so that psychiatrists may concentrate their prescribing efforts on difficult-to-manage patients who may be treatment-resistant and/or non-compliant and those patients who have substantial physical health co-morbidity. Another advantage might be that CPNs, who are case managers, may be able to spend more time than their psychiatrist colleagues in the detailed evaluation of effectiveness and side effect monitoring and management.

Cognitive behaviour therapy

For more than 35 years, nurses in the United Kingdom have been trained to provide psychological treatment to patients with various mental health problems. These developments began in 1972, when Isaac Marks, a psychiatrist working at the Maudsley hospital, began a 3-year experiment to determine whether nurses could be trained to deliver behavioural interventions for neurotic disorder. Isaac Marks was one of the first to recognize that the workforce of psychologists would be insufficient to deliver evidence-based treatment. Subsequently, Marks[19] published data which demonstrated both the clinical and economic effectiveness of nurses working with neurotic disorders in primary care. Over the years, training programmes for nurses have developed and now nurses are trained in a variety of university and clinical settings alongside their psychology colleagues in the practice of evidence-based psychological treatments, i.e. cognitive behaviour therapy for a very wide range of disorders. Whilst the original efforts to train nurses were centred on techniques for the treatment of phobias and obsessive–compulsive disorder, nurses are now trained more comprehensively in cognitive behavioural methods, which encompass treatment techniques used in the treatment of depression, schizophrenia, and personality disorders. In recent years there has been a small, but significant, growth in the United Kingdom of psychiatric nurses employed as psychological therapists. However, this trend has not been replicated in Australasia or Europe, where legislative frameworks prevent nurses from obtaining full accreditation as psychological therapists. In the United States, the situation is variable. Nevertheless, the American Psychiatric Nurses' Association membership comprises a significant number of nurses who have full accreditation as therapists in their respective states. Such nurses are now often prepared at a post-doctoral level and their expertise is arguably equivalent to that of their clinical psychologist colleagues.

Primary care

Following the refocus of CPNs efforts on people with schizophrenia and other serious and enduring illnesses, there have recently been a number of United Kingdom policy developments that will lead to more psychiatric nurses being employed in primary care settings. This trend follows the recognition that many people with common mental health problems do not receive evidence-based treatment or, if they do, they are subject to long periods on waiting lists. Psychiatric nurses are now being trained to provide brief evidence-based interventions in primary care and also to provide important assessment and screening functions, so as to ensure that patients who need the services of the community mental health teams are suitably referred and those who can be managed at primary care level are provided with appropriate treatments. Nurses are now also increasingly involved in the delivery of computerized cognitive behaviour therapy, which, as a recent NICE review demonstrated,[20] is effective in the treatment of a wide range of mental health problems. This method of treatment is particularly important given the scarcity of skilled therapist resources. Psychiatric nurses are therefore increasingly involved in the running of 'Computer clinics', which now use a very wide variety of treatment packages for a whole range of disorders. Whist most of these packages come at a cost, there are now free to access programmes on the Internet. These include 'Moodgym',[21] a programme for the self-help treatment of depression, which is based on a cognitive behavioural approach and developed in Australia at the Australian National University in Canberra. Moodgym has been evaluated and its use is supported by positive randomized trial data.[22]

Psychiatric nursing in the developing world

As noted above, there have been a number of training initiatives in the former communist countries of Eastern Europe and the former Soviet Union.[1] Nevertheless, the numbers of trained psychiatric nurses across this region still remains fairly small. It is also clear that the skills of psychiatric nurses, in these countries, are compromised by the relatively poor general standards of mental health care.

With the exception of South Africa, psychiatric nursing is very poorly developed in the African continent. However, the World Health Organization has a number of projects that aim to integrate mental health into primary care provision. For example, the Ministry of Health in Kenya, the Kenyan Psychiatric Association, and the Kenyan Nursing Council are working with the UK Department for International Development on a 5-year programme

that began in 2005, which has the ambitious aim of training 3000 primary care workers with some skills in very basic mental health care.

In Asia, psychiatric nursing is becoming increasingly recognized and there are now substantial training initiatives in the Indian subcontinent and China. Nevertheless, the numbers of psychiatric nurses who are needed in those vast countries are very great and, at present, the best way of describing the psychiatric nursing presence would be to say that it is sparse and patchy, with most nurses working in the large cities. An additional problem is that, where reasonable standards of education and training exist (and this applies particularly to India and China) there is a considerable loss to immigration to the developed world—to countries such as the United Kingdom and the United States, where the recruitment and retention of psychiatric nurses is an ongoing problem, and is financially very attractive.

There is obviously enormous potential for the development of psychiatric nursing across the developing world and, although current initiatives have focused on providing nurses with basic skills, there is obviously enormous potential for the provision of evidence-based psychosocial interventions and, given the tremendous shortage of psychiatrists in many countries, nurse prescribing (providing that the nurses have received adequate education and training) could obviously potentially provide widespread benefits for a substantial proportion of the literally millions of people whose illnesses are currently untreated.

Conclusion

Psychiatric nursing is a relatively new profession, which has evolved over the past 50 years, from a branch of general nursing, where the main role focused on the custodial care of people with chronic serious and enduring mental illnesses—such as schizophrenia—to the present day situation, where nurses are, in the developing world at least, more likely to be university graduates who may be employed in a number of very diverse roles. The setting for these roles could be in inpatient settings, where custodial care is challenging, to say the least, to the community with roles involving psychosocial interventions within primary health care teams, to inpatient and community settings where nurses are increasingly autonomous prescribers of medication or psychological therapists. It is pleasing to note that the education and training of psychiatric nurses is gradually becoming more evidence based and policy makers are apparently much more aware of the need to provide focused skill sets on populations of need.

Further information

From values to action: The Chief Nursing Officer's review of mental health nursing (2006)—England—http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4133839

American Psychiatric Nurses Association—http://www.apna.org/i4a/pages/index.cfm?pageid=1

Newell, R. and Gournay, K. (2008). *Mental health nursing: an evidence based approach* (2nd edn). Elsevier, London.

Australian and New Zealand College of Mental Health Nursing—http://www.acmhn.org/index.html

References

1. Geneva Initiative Training. http://www.geneva-initiative.org/pages/projects/projects.asp
2. Healthcare Commission. (2005). The National Audit of Violence 2003–2005. www.healthcarecommission.org.uk
3. National Confidential Inquiry. (2006). *Avoidable deaths: five year report of the National Confidential Inquiry into suicide and homicide by people with a mental illness*. University of Manchester. www.medicine.manchester.ac.uk/suicideprevention/nci/useful/avoidable_deaths.pdf
4. National Institute for Health and Clinical Excellence (NICE) (2005). Violence. The short-term management of disturbed/violent behaviour in psychiatric inpatient settings and emergency departments. *Clinical Guidelines 25*. www.nice.org.uk
5. Drury, V., Birchwood, M., Cochrane, R., *et al.* (1996). Cognitive therapy in recovery from acute psychosis, a controlled trial. *The British Journal of Psychiatry*, **169**, 593–607.
6. Kemp, R., David, A., Hayward, P., *et al.* (1998). Compliance therapy, an 18 month follow up. *The British Journal of Psychiatry*, **5**, 228–35.
7. Sailas, E. and Fenton, M. (2002). Seclusion and restraint for people with serious mental illnesses. *The Cochrane Library*, (1). Update Software, Oxford.
8. Paykel, E., Mangen, S., Griffith, J., *et al.* (1982). Community psychiatric nursing for neurotic patients: a controlled trial. *The British Journal of Psychiatry*, **140**, 573–81.
9. White, E. (1990). *The third quinnenial survey of CPNs*. Department of Nursing Studies, University of Manchester.
10. Gournay, K. and Brooking, J. (1994). The CPN in primary care: an outcome study. *The British Journal of Psychiatry*, **165**, 231–8.
11. Gournay, K. and Brooking, J. (1995). The CPN in primary care: an economic analysis. *Journal of Advanced Nursing*, **22**, 769–78.
12. Brooker, C., Fallon, I., Butterworth, A., *et al.* (1994). The outcome of training community psychiatric nurses to deliver psychosocial intervention. *The British Journal of Psychiatry*, **165**, 222–30.
13. Burns, T., Fiander, M., Kent, A., *et al.* (2000). Effects of caseload size on the process of care of patients with severe psychotic illness: report from the UK700 trial. *The British Journal of Psychiatry*, **177**, 427–33.
14. Gray, R., Wykes, T., Edmonds, M., *et al.* (2004). Effect of a medication management training package for users on clinical outcomes for patients with schizophrenia: cluster randomised controlled trial. *The British Journal of Psychiatry*, **185**, 157–82.
15. Gray, R., Leese, M., Bindman, J., *et al.* (2006). Adherence therapy for people with schizophrenia. *The British Journal of Psychiatry*, **189**, 508–14.
16. American Psychiatric Nurses' Association. www.apna.org
17. Extended Nurse Prescribing. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH 4006775
18. Gournay, K. and Gray, R. (2001). *Should mental health nurses prescribe? Maudsley discussion paper*. Institute of Psychiatry, London.
19. Marks, I. (1985). *Nurse therapists in primary care*. RCN Publications, London.
20. National Institute for Health and Clinical Excellence (NICE) (2006). The clinical and cost effectiveness of computerised cognitive behaviour therapy for depression and anxiety. *Technology Appraisal 51*. NICE, London. http//www.nice.org.uk/
21. Moodgym website: http://moodgym.anu.edu.au/
22. Christensen, H., Griffiths, K.M., and Jorm, K. (2004). Delivering interventions for depression by using the internet: randomised controlled trial. *British Medical Journal*, **328**, 265.

6.4.3 Social work approaches to mental health work: international trends

Shulamit Ramon

The historical development of mental health social work

Social work was formally established in most European countries and North America at the end of the nineteenth century, before it took off gradually in other countries. It usually developed out of charitable work, which focused on financial support for poor families. The second main strand in social work was represented by the Settlement Movement, which concentrated on improving the communal life of poor people by living with them, using community work methods to support and empower.

The major impetus to developing mental health social work at the beginning of the twentieth century was related to the work of leading psychiatrists and psychologists with shell-shocked (PTSD) soldiers during the First World War.[1] This approach lead to the establishment of the psychodynamically oriented Tavistock Clinic in London, where the first British psychiatric social worker was appointed in 1920.[2] Stuart[3] argues that American mental health social work in the pre-1920 period concentrated more on care in the community in its social, rather than its administrative or psychological meaning, than after 1920, when it shifted further to the psychological dimension.

Social workers in both the children and the adults outpatient services provided comprehensive psychosocial history of the child/adult and their family, enabled parents, teachers, and partners of adult clients to understand the underlying psychological reasons for the index client's mental ill health, and guided them as to how they could actively support that family members.[2] It was only in the 1950s that qualified psychiatric social workers began to work in hospitals.

Since the 1970s all English-speaking countries have also opted for deinstitutionalization as their core mental health policy, leading to the closure of many of their psychiatric hospitals, replacing them by community-based services and by small psychiatric wards in general hospitals.[4,5] With the notable exception of Italy,[6,7] most continental European countries have opted for bed reduction coupled with less extensive community services.

The trend towards deinstitutionalization is formally adhered to also in Latin America, but thus far is only practised in some small-scale projects.[8,9] This applies also to Asia. MHSWs (mental health social workers) there focus mainly on sorting out benefits, though some are based in rehabilitation focused facilities where they work on connecting users to educational and employment opportunities.[10]

This fundamental change has led to the relocation of MHSWs away from institutions into community services,[5] to a renewed interest in rehabilitation, and more recently also in the newly defined recovery.[11]

The paralleled development of private, for-profit mental health services especially—but not exclusively—in the United States, has led to a further shift in the location of MHSWs and their work focus. Most United States MHSWs are to be found today working as psychotherapists in private practice or in managed care residential units.[12] In the latter they work with users who have long-term mental illness to a per capita budget.

Although the not-for-profit sector has grown considerably with the focus on care in the community, MHSWs work there only in certain countries in which the public sector has either been reduced or never played the major part it does in the United Kingdom (e.g. the Netherlands, Hong Kong).

Underpinning values

Social work, including its mental health branch, is ethically governed by a set of values, which are expected to be universal and adhered to in everyday practice,[13] even though its implementation may prove at times to be problematic in terms of balancing care and control.

The values are derived from the liberal collectivist, humanistic, tradition of the twentieth century in which social work has developed.

The core values are social justice, respect for people who social workers meet at their most vulnerable state, readiness to help in a way which will enable the client to retain dignity, self-determination, and enhance their problem-solving abilities. Social workers are expected to take an active stance against any type of discrimination. Furthermore, social workers are committed to pursing a psychosocial approach in any type of their practice, and believe that most clients have the potential to grow and positively change.

Several elements stand out as central to MHSW:

1 The right to fail—this comes as part of the right to self-determination, in that social workers are aware that risk needs to be taken at times to enable people to grow and develop, or as a basic human right of making a mistake. When social workers take this right seriously, they are able to have a genuine discussion with clients as to the pros and cons of risk-taking, of learning from success as much as of learning from failure.[14,15]

2 The wish to take an active stance against discrimination applies to working well with clients who come from ethnic minorities, from sexual orientation minorities, and to combating stigma against mental illness in one's practice.

3 The adherence to a psychosocial approach entails ensuring that both the psychological and the social aspects of users' lives are attended to, an issue of importance in mental health where often biological aspects are attended to, but the psychosocial ones are not getting the same priority.[16]

Conceptual developments

Psychodynamic approaches

As outlined above, mental health social work originated within the psychodynamic fold, though social workers did not practice psychoanalysis as a work method.

Social workers have tended to select from the range of psychodynamic perspectives those theories, which were more focused on the ego, rather than on the id or the unconscious. The impact of ego psychology was/is in evidence in terms of understanding how people come to develop and maintain mental distress and mental

illness, the importance of family dynamics, and of attachment to significant others.[17, 18, 19, 20]

American social workers developed the crisis approach in its application to all areas of social work.[21] Based on Erickson's notion of the normal crisis every person goes through when moving from one stage of life to another, major life events may lead initially to adverse reactions. However, with professional support people can reorganize their reactions more constructively, reduce the duration of these reactions, be more ready for change at the point of crisis, and learn how to improve their coping strategies and emotional responses. The problem-solving approach also originated from the United States, developed by Perlman.[22] Although the psychodynamic understanding of relationships is in evidence in her work, she focused on the process of social work with individuals and families (casework) and the client–worker relationships, beginning with the presenting problem.

The identification of child abuse, especially child sexual abuse, and its implications for the mental health of children and adults in the 1970s and the 1980s led to refocusing on the psychodynamic approach among social workers in this area[23] at a time in which all other approaches have paid less attention to the impact of such abuse on mental health.

Learning theory applications in social work

Behavioural social work

Behavioural social work developed in the United States in the 1950s, and is a leading approach in relation to people with milder forms of mental illness and problems of living.[24] Its application within social work does not differ in any significant way from its application within psychiatry or psychology. In this sense it is not a social work approach. A number of influential texts appeared in the United Kingdom which demonstrated the research evidence pertaining to the effectiveness of the approach in a number of social work areas.[25]

Task-centred social work

This orientation takes further the crisis perspective and the lessons from learning theories and behaviour modification.

Reid and Epstein[26, 27], as well as Marsh and Doel,[28] proposed that people work better on their problems if focused on specific targets and if the problem-solving effort leads to success, however small. Research evidence demonstrated the usefulness of this approach to different aspects of social work, such as direct work with children and their parents, as well as with people suffering from mild mental distress symptoms.

The social dimension in mental health social work

Social workers and theorists interested in the social dimension began usually from the assumption that inequality in opportunities and in civic participation due to poverty may increase the rate of mental illness among poor people. This assumption follows Merton's classical matrix of the reactions to the gap between social goals and means, in which mental illness is a reaction of people who accept socially desirable goals, but withdraw from obtaining them after being frustrated in doing so, whilst at the same time not adopting antisocial means (as in criminal behaviour) or developing an alternative model of society (social rebels).

This strand of thinking was reinforced in the 1960s and 1970s by the application of Marxist thinking and the combined impact of the deviancy and anti-psychiatry orientations.[29] Discrimination on the basis of age, ethnicity, gender, or sexual orientation was added in the 1980s to the likely social factors which foster inequality.

Social workers accepted the logic presented by sociologists such as Goffman and Scheff[30, 31, 32] that the stigma attached to mental illness is largely irreversible, as it is accepted both by others and by the individual concerned who in turn internalizes his or her poor social status.

Interestingly, although accepting the enormity of the labelling process, social workers did not count themselves among the labellers.

The appeal of the anti-psychiatry approach for social workers related to acknowledging the price of labelling for the individual concerned, and the considerable shortcomings of a system focused on the psychiatric hospital and medication in which psychological and social factors were largely ignored.

Today the social perspective implies a greater focus on social inclusion, supporting users and carers-led initiatives, ensuring financial support side-by-side with the critique of the medicalized approach to mental health and illness, of modernity and post-modernity.[16]

A more recent strand of this approach is outlined in the critical social work, which applies a post-modern perspective to the analysis of where social work is, as well as the issues and dilemmas related to mental health social work. Bainbridge[33] highlights the need to focus on the social dimension and sociological understanding of mental ill health in social work, as well as on issues of power and empowerment.

(a) The social role valourization (SRV) and the strength approach

This approach was initially developed by psychologists in the field of learning difficulties.[34–37]

SRV accepts the deviancy approach up to the point at which the impact of labelling and segregation is said to be irreversible. Conversely, SRV is focused on reversing the devaluation of the disabled person and the group, while accepting that a disability exists. The devalued existence can be reversed by the combined impact of the following:

- enabling those who have been segregated to live in the community by providing them with the opportunities to do so and the support they require for this purpose
- enhancing the competencies of the disabled person
- changing their public image, in part by their positive presence in ordinary settings in the community
- upgrading the state of the physical settings in which a disabled group is treated, lives, and works
- changing the derogatory language used in both professional and lay circles in describing people with disabilities.

Its protagonists are critical of professional attitudes, knowledge, and skills, including those of social workers (see Wolfensberger).[35] Yet as a group social workers have within their repertoire more of the attitudes, knowledge, and skills required by this approach than any other mental health profession. Furthermore, SRV offers an

interesting and comprehensive combination of psychological and social dimensions; for an application to how it can work with the Nearest Relative in mental health.[38]

In the United States and Canada, but much less so in the United Kingdom, SRV came into prominence within social work through the **strengths** model of social work.[39] The model is unique in concentrating on the strengths the person and his or her environment possess, and how these could be harnessed to solve the specific problem and lead to an improvement in the person's quality of life. Coming together with a focus on following people's ambitions (as long as these are within socially acceptable norms), this orientation has led to useful and positive outcomes in care management.[40] A further development of the strengths model is the growing interest in focusing on enhancing resilience in mental health social work.[41]

Interestingly, the approach has been adopted by other mental health professionals in the field of employment without acknowledging the debt to social work.

Legally anchored MHSW

The social mandate of social work is anchored within legal and policy frameworks; this applies to MHSW too.

Securing benefits

In all countries social workers are gatekeepers to and advocates for securing benefits either in cash (e.g. disability allowance, Direct Payment agreement) or in kind (e.g. housing, clothing). They often have to make the claim in addition to the client, verify the claim as against eligibility criteria, secure supporting documents from other professionals, at times negotiate with other agencies (e.g. social security, health, education), and in a number of countries they are indeed located in social security services (e.g. Portugal, Israel).

While this work is considered to be routine, its importance in the life of poor people cannot be underestimated. The evidence highlights that most people with enduring mental illness are poor,[42] and that remaining poor is a strong counter-indication to becoming mentally healthy. Furthermore, the evidence related to Direct Payment in mental health,[43] which enables users to take the driving seat as to how they spend their budget in agreement with the local authority and mental health trust, illustrates the social inclusion and recovery value of such a scheme which clearly comes out of the strengths model.

Knowledge of the available resources and eligibility is needed for this type of work, as well as the ability to inform users and enable them to participate as deserving partners.

In a large number of countries this is the main social work task in mental health (e.g. Brazil, Greece, Italy, Ireland, Portugal, and Poland).

The approved social worker

This role illustrates an enhanced legal position for MHSWs; its fullest form is practised in the United Kingdom.

The approved social worker was developed in Britain to provide a complementary measure to the psychiatric perspective within the 1982 amendments to the 1959 Mental Health Act.

Social workers were seen as suitable professional figures who would represent the psychosocial angle in parallel to the psychiatric view in the following instances:

- assessing people when an application has been made for a compulsory admission to a psychiatric unit
- the follow-up to such an admission
- mental health review tribunals (established within the 1959 Mental Health Act)
- work with the Nearest Relative[38]
- coordinate the multi-disciplinary assessment, which needs to be carried out by a psychiatrist and a GP in addition to the social worker. This assessment has to be carried out within a specified limited period of time.

Each of these tasks calls for somewhat different knowledge and skills, as well as emphasis and use of a range of more generic skills.[44] In each task social workers are asked not to replicate the psychiatric assessment but to compliment it. For example, they have to look for the least restrictive alternative to the hospitalization before they can recommend a hospital admission, rather than diagnose mental illness. Social workers have an autonomous position as they can disagree with the views of the other professions. The role requires exercising more social control than care, a contested issue within social work.

Training to become an approved social worker requires 60 days of academic input and supervised practice initially, followed by 5 days refresher training annually. Individual social workers can take it up after 2 years of post-qualification work experience. This compares with 2 days training for general practitioners and 1 day for psychiatrists.

Most of the activities undertaken by MHSWs with adults since 1983 are related to meeting the requirements of this role. Existing evidence[45] highlights that in most cases ASWs are working to a good standard. The role also offered MHSW a higher status and pay, but came with the price tag of giving up most of their previous activities, such as family work, group and community work, and work with users who have minor mental illness. However, the proposed new English Mental Health Act[46] includes the introduction of AMHPs (Approved Mental Health Practitioners) who can come from any mental health discipline, but likely to be nursing because of the numerical dominance of nurses in the English mental health service. ASWs are unhappy at being dethroned of their unique legal role, arguing that nurses do not have the same background training for psychosocial understanding and intervention.

Workforce research into ASWs[47] has highlighted a recent steady decrease in the number of ASWs, a high number of workers approaching retirement age and low morale, factors likely to have played a part in the government's wish to introduce other professions to this role. There are currently 4500 ASWs, a minority in the total workforce of social workers in England which stands at 46 000.

While current ASWs and MHSWs will be able to be part of the AMHP workforce, this change may also enable them to reclaim some of their previous roles and activities.

Care management

In all English-speaking countries and a number of European countries (e.g. the Netherlands, Slovenia), MHSWs are also often engaged in one form or another of care management, which follow the specific laws and regulations of each country (e.g. the Care in the Community 1990 Act in England).

Care management (not be confused with managed care) is a form of coordinating the assessment, planning, and interventions with people who require long-term care, including in mental health. It is aimed at preventing fragmentation and duplication of professional input and services, as well as ensuring that services follow the user's needs, and not vice versa as was—and still is—the case all too often. While in the United Kingdom, psychiatrists are formally the nominated care manager since 1995, and CPNs are the professionals in more frequent contact with the users, in other countries such as Australia, New Zealand, and Canada, social workers are often the nominated care coordinator.

Care management can be practised in a variety of ways, ranging from a purely administrative orientation, through clinical care management, to one anchored within the strengths and recovery orientations.[40, 48, 49] The choice is often dictated by the managers of a local authority or a mental health trust.

Community treatment orders (CTOs)

CTOs constitute a third legally defined area in which MHSWs are engaged. They originated in the United States, and exist in a variety of formats in Canada, Australia, and the United Kingdom.[50–52] They represent a response to the escalation of concerns about risk avoidance in the field of mental health, closely related to the growing fears of risk raised by modernity and post-modernity,[54, 55] and reinforced more recently by fears of terrorism as a particularly threatening type of risk. It is indicative that this preoccupation is not shared between North and South Europe; it is much more prominent in the North.

Following a legal process, CTOs enable the nominated mental health professional to require a service user to adhere to specific restrictions, such as to live in a certain facility or to present themselves for interventions at specific locations. This measure has been introduced to ensure that users with long-term mental illness who lead a disorganized life (e.g. often do not comply with medication, live rough, misuse substances, mishandle money, misagreed appointments, and get into trouble with the law) will have a safety net, which structures their lives. There is some evidence of the effectiveness of CTOs,[48, 49] but it seems restricted to specific subgroups within the broader category of people with severe and enduring mental illness.

Social workers are often the nominated professionals responsible for the agreed plan and enforcement of the CTO. This puts them in a somewhat conflictual position regarding the desired focus on care in social work, as it is tilted more towards the control element.[56]

Non-legally anchored MHSW

The constraints on this type of work come not only from the primacy of legally sanctioned work, but also from working within welfare bureaucracies, for a private employer interested primarily in profit, or for an impoverished not-for-profit service.

Despite these constraints, we have examples of good and innovative MHSW practice in most countries, which include:

- Successful attachment to primary care was established as early as 1965 in London, with the social workers, their clients, and the general practitioners expressing satisfaction with this way of working. Nevertheless, this form was largely abandoned owing to the focus on statutory responsibilities.
- Social workers pioneered collective user involvement approach during the early 1980s[57] some years before it became fashionable in wider circles.
- Applying self-directed group and community work approach to working with mothers of abused children, empowering them to take control over their lives.[58]
- Initiating de-institutionalization in social care institutions in Slovenia.[59]
- Creating family support teams[60] where most staff members are social workers providing brief assessment and consultation to families of children with minor mental distress, and are often successful in preventing the need for referral to more expensive services.
- Establishing the Building Bridges project in which parents with mental health difficulties and their children are supported together as well as separately.[61]
- Creating The Faith Links project, a multi-faith project within an inpatient service in which users, volunteers, and social workers plan joint activities which follow users' spiritual wishes.[62]
- Supported education and employment schemes in mental health initiated by social workers.[63]

Conclusion

Mental health social work is a broad, rather than a rigorous, church. Since the 1980s social workers have gained in professional status by the introduction of the roles of the approved social worker (or licensed to carry out civil commitment in the American context), care co-ordinators, managers of managed care facilities, or psychotherapists. These gains have come at a price outlined in the text above.

Often the cost of closer collaboration within the multi-disciplinary framework has led to the risk of giving up the attempt to hold on to, and further develop, an alternative and complimentary perspective from that of psychiatrists, nurses, or psychologists, as well as raising doubts as to the uniqueness of MHSW.

The increased narrowness of the role is not simply the byproduct of the legal framework. It is also due to increased specialization within mental health on the one hand, and the effects of neo-liberal policies globally on public sector funding on the other hand.

The move to privately contracted work, either in managed care or in psychotherapy so apparent in the United States, is yet another outcome of neo-liberal policies which fragments MHSW. As a trend we are likely to see growing beyond the United States, the increased concentration of mental health social workers within the private sector does not bode well for a profession whose value base focuses on the need to protect the more vulnerable and stigmatized populations, and to provide the dual perspectives of psychosocial input.

Mainly due to governmental pressure related to fear of risk and its potential political fallout, the focus on working exclusively with people experiencing long-term severe mental illness has contributed to the increasing narrowness of the role of social workers in most First World countries. The paralleled withdrawal of social work involvement with people who have milder forms of mental distress within public sector and not-for-profit services, and its

increased availability only to those who can afford it, is a reflection of this situation.

The core qualities of belief, optimism, and caring of MHSWs identified in a cross-national research[64] coupled with the ability of MHSW to innovate as highlighted in this chapter, illustrate the optimistic scenario for positive change within this branch of social work. However, unless theory building and research aspects are given the importance they deserve within MHSW globally, including an inevitable critical dimension of the existing system, mental health social work is likely to be no more than a reflection of the developments in other professions. This will not only mean curtailing its autonomous potential, but also the impoverishment of the multi-disciplinary framework as a whole of a crucial dimension necessary for its comprehensive work, as exemplified in some recent work on the social aspects of MHSW.[16]

In addition, mental health social work will have to develop a much stronger policy making function, if it is to provide a more responsive, effective, and comprehensive service to users, relatives, and the communities in which these people live.

Further information

Norman, E. (ed.) (2000). *Resiliency enhancement: putting the strength perspective into social work practice*. Columbia University Press, New York.

Ramon, S. and Williams, J.E. (eds.) (2005). *Mental health at the crossroads: the promise of the psychosocial approach*. Ashgate Publishing, Aldershot.

Tew, J. (ed.) (2005). *Social perspectives of mental health*. Jessica Kingsley, London.

References

1. Dicks, H. (1970). *Fifty years of the Tavistock*. Tavistock Publications, London.
2. Timms, N. (1964). *Social casework*. Routledge and Kegan Paul, London.
3. Stuart, P.H. (1997). Community care and the origins of psychiatric social work. In *Social work in mental health: trends and issues* (ed. U. Aviram), pp. 25–37. Haworth Press, New York.
4. Goodwin, S. (1997). *Comparative mental health policy: from institutional to community care*. Sage, London.
5. Shera, W., Aviram, U., Healy, B., *et al.* (2002). Mental health systems reform: a multi country comparison. *Social Work in Health Care*, **35**, 547–75.
6. De Leonardis, O., Mauri, D., and Rotelli, F. (1986). *Deinstitutionalisation: a different path: the Italian mental health reform, health promotion*, Vol. **2**, pp. 151–65. WHO Cambridge University Press.
7. Ramon, S. (ed.) (1990). *Psychiatry in transition: British and Italian experiences*. Pluto Press, London.
8. Vasconcelos, E.M. (2005). Structural issues underpinning mental health care and psychosocial approaches in developing countries: the Brazilian case. In *Mental health at the crossroads: the promise of the psychosocial approach* (eds. S. Ramon and J.E. Williams), pp. 95–108. Ashgate Publishing, Aldershot.
9. Ramon, S. and Williams, J.E. (2005). *Mental health at the crossroads: the promise of the psychosocial approach*. Ashgate Publishing, Aldershot.
10. Fung Sheung Chee, B., Law Ka Sin, J., and Lee Yuk Yee, K. (2006). *Clubhouse model in an Asian culture, 5th International Conference on health and mental health social work*, December, Hong Kong.
11. Ramon, S., Healy, B., and Renouf, N., (2007). Recovery from mental illness as an emergent concept and practice in Australia and the UK. *International Journal of Social Psychiatry*, **53**, 108–22.
12. Cohen, G.A. (2003). Managed care and the evolving role of the clinical social worker in mental health. *Social Work*, **48**, 34–43.
13. Beckett, C. and Maynard, A. (2005). *Values and ethics in social work*. Sage, London.
14. McDermott, R. (ed.) (1975). *Self determination in social work*. Routledge and Kegan Paul, London.
15. Ramon, S. (2006). Risk avoidance and risk taking in mental health social work. In *Knowledge in mental health: reclaiming the social* (eds. L. Sapouna and P. Hermann), pp. 39–56. Nova Publications, New York.
16. Tew, J. (ed.) (2005). *Social perspectives of mental health*. Jessica Kingsley, London.
17. Parad, H.J. (ed.) (1958). *Ego psychology and dynamic casework*. American Family Services Association, New York.
18. Hutton, J.M. (ed.) (1977). *Short-term contracts in social work*. Routledge and Kegan Paul, London.
19. Yellowly, M. (1980). *Psychoanalysis and social work*. Van Nostrand, New York.
20. Howe, D. (1995). *Attachment theory for social work practice*. Macmillan, London.
21. Golan, N. (1978). *Treatment in crisis situations*. Free Press, New York.
22. Perlman, H. (1957). *Social casework—a problem-solving process*. Chigaco University Press, Chigaco, IL.
23. Perlberg, R. and Miller, A. (eds.) (1992). *Gender and power in families*. Routledge, London.
24. Gambrill, E.D. (1977). *Behaviour modification*. Jossey-Bass, San Francisco, CA.
25. Hudson, B. and Macdonald, E. (1986). *Behavioural social work: an introduction*. Macmillan, London.
26. Reid, W.J. and Epstein, L. (1972). *Task centered casework*. Columbia University Press, New York.
27. Jackson, V.H. (1996). Behavioral managed care: a social work perspective. *Behavioral Health Management*, 22–3.
28. Marsh, P. and Doel, M. (1993). *Task-centred social work*. Ashgate, Aldershot.
29. Leonard, P. and Corrigan, P. (1978). *The Marxist approach to social work*. Macmillan, London.
30. Goffman, I. (1961). *Asylums*. Penguin, Harmondsworth.
31. Laing, R.D. (1965). *The divided self: an existential study in sanity and madness*. Penguin, Harmondsworth.
32. Scheff, T. (ed.) (1975). *Labelling madness*. Prentice-Hall, Englewood Cliffs, NJ.
33. Bainbridge, L. (1999). Competing paradigms in mental health practice and education. In *Transforming social work practice: postmodern critical perspective* (eds. B. Pease and J. Fook), pp. 179–94. Routledge, London.
34. Nirje, B. (1969). The normalisation principle and its human management implications. In *Changing patterns in residential services for the mentally retarded* (eds. R. Kugel and W. Wolfensberger), pp. 255–87. President's Committee on Mental Retardation, Washington, DC.
35. Wolfensberger, W. (1983). Social role valorisation: a proposed new term for the principle of normalisation. *Journal of Mental Retardation*, **21**, 234–9.
36. Ramon, S. (ed.) (1991). *Beyond community care: normalisation and integration work*. Mind/Macmillan, Basingstoke.
37. Brandon, D. (1991). Implications of normalisation work for professional skills. In *Beyond community care: normalisation and integration work* (ed. S. Ramon), pp. 35–55. Mind/Macmillan, London.
38. Rapaport, J. (2005). The informal caring experience: issues and dilemmas. In *Mental health at the crossroads: the promise of the psychosocial approach* (eds. S. Ramon and J. Williams), pp. 155–70. Ashgate Publishing, Aldershot.
39. Saleebey, D. (ed.) (1992). *The strengths perspective in social work practice*. Longman, New York.

40. Rapp, C. (1998). *The strengths perspective of case management with persons suffering from severe mental illness*. Oxford University Press, Oxford.
41. Norman, E. (ed.) (2000). *Resiliency enhancement: putting the strengths approach into social work practice*. Columbia University Press, New York.
42. Pilgrim, D. and Rogers, A. (2003). *Mental health and inequality*. Palgrave, Macmillan, Basingstoke.
43. Glasby, J. and Lester, H. (2002). *Social work and direct payment*. Policy Press, Bristol.
44. Barnes, M., Bowl, R., and Fisher, M. (1990). *Sectioned: social services and the 1983 Mental Health Act*. Routledge, London.
45. Hatfield, B. and Robinshaw, P. (1994). The use of compulsory powers by approved social workers in five local authorities: some trends over two years. *Journal of Mental Health*, **3**, 339–50.
46. Rapaport, J. (2006). New roles in mental health: the creation of the approved mental health practitioner. *Journal of Integrated Care*, **14**, 37–46.
47. Huxley, P., Evans, S., Webber, M., *et al.* (2005). Staff shortages in the mental health workforce: the case of the disappearing social worker. *Health and Social Care in the Community*, **13**, 504–13.
48. Kanter, J. (1989). Clinical case management: definition, principles, components. *Hospital and Community Psychiatry*, **40**, 361–8.
49. Brandon, D., Atherton, K., and Brandon, A. (1996). *Handbook of care planning*. Positive Publications, London and ref. 26.
50. Hiday, V.A. and Scheid-cook, T.L. (1991). Outpatient commitment for "revolving door" patients compliance and treatment. *Journal of Nervous and Mental Disease*, **179**, 83–8.
51. Campbell, J., Brohpy, L., Healy, B., *et al.* (2006). International perspectives on the use of community treatment orders: implications of mental health social workers. *British Journal of Social Work*, **36**, 1101–18.
52 Canvin, K., Bartlett, A., and Pinfold, V. (2002). A "bittersweet pill to swallow": learning from mental health service users' responses to compulsory community care in England. *Health and Social Care in the Community*, **10**, 361–9.
53. Stanley, N. and Manthorpe, J. (2004). *The age of the inquiry*. Routledge, London.
54. Beck, A. (1992). *The risk society*. Sage, London.
55. Rose, N. (2002). *Powers of freedom: reframing political thought*. Cambridge University Press, Cambridge.
56. Thompson, P. (2003). Devils and deep blue seas: the social worker in-between. *Journal of Social Work Practice*, **17**, 35–47.
57. Hennelly, R. (1990). Mental health resource centres. In *Psychiatry in transition: British and Italian experiences* (ed. S. Ramon), pp. 208–18. Pluto Press, London.
58. Mullender, A. and Ward, D. (1991). *Self-directed groupwork*. Whiting and Birch, London.
59. Flaker, V., Cizely, M., Ferle, Z., *et al.* (2004). Special care homes: the vision: a project of the community of social institutions in Slovenia. *Journal of Social Work, Faculty of Social work*, University of Ljubljana, Ljubljana, Slovenia.
60. Debell, D. and Walker, S. (2003). *Norfolk family support teams: an evaluation of the first two years*. Anglia Ruskin University, Cambridge.
61. Diggins, M. (2000). Innovation as a professional way of life—the building bridges project for parents-users of mental health services and their children. In *A stakeholder approach to innovation in mental health services* (ed. S. Ramon), pp. 77–93. Pavilion, Brighton.
62. Jones, J. (2006). *The faith links project—equality and diversity in Brent, central and north west London*. Mental Health NHS Trust, London.
63. Mowbray, C.T., Collins, M.E., and Bellamy, C.D. (2005). Supported education for adults with psychiatric disabilities: an innovation for social work and psychosocial rehabilitation practice. *Social Work*, **50**, 7–20.
64. Ryan, M., Merighi, J.R., Healy, B., *et al.* (2004). Belief, optimism and caring: findings from across-national study of expertise in mental health social work. *Qualitative Social Work*, **3**, 411–29.

6.4.4 Art therapy

Diane Waller

The fundamental principles of art therapy/art psychotherapy

Definitions

Descriptions of art therapy from two of the oldest and largest professional associations, the British Association of Art Therapists and the American Art Therapy Association refer to: the use of art materials for self-expression and reflection in the presence of a trained art therapist. Art therapy uses the flexible, creative problem-solving potential of art-making to improve and enhance the physical, mental, and emotional well-being of individuals of all ages. The relationship between the therapist, client, and their artwork is of central importance. Art therapy can be used on a one-to-one and group basis.

Art therapy (or art psychotherapy, both titles are protected by law) in the United Kingdom is firmly rooted in psychodynamic and humanistic concepts and practices appropriate to public sector settings, and adapted to the social and mental health of the client. It is a broad-based discipline, involving substantial knowledge of the visual arts, individual and group psychotherapy, social and communication sciences, and the impact of culture on health.[1]

Main premises

- That visual image-making is an important aspect of the human learning process;
- That art made in the presence of an art therapist may enable a person to get in touch with feelings that cannot easily be expressed in words;
- That the creative process helps people to resolve conflicts and problems;
- That art can act as a 'container' for powerful emotions and be a means of communication between client(s) and therapist;
- That the image can serve to illuminate the transference in the case of a psychodynamic approach.

Engagement in image-making is of central importance although clients do not need any prior experience of or skill in art, as the aim is not to produce a 'good' piece of art that can be exhibited. The images made in art therapy may embody thoughts and feelings, be a bridge between the 'inner world' and outer reality, be a mediator between unconscious and conscious, hold and symbolize past, present, and future aspects of a client's life. Ambivalence and conflict can be stated and contained within an image. In art therapy the client tries to give form to what seem to be inexpressible or unspeakable feelings, which they can then share with the art therapist.

The focus of the transference (bringing feelings from the past into the present), can be onto the art object rather than to the therapist directly, adding a 'third dimension' to the therapeutic process.[2–4]

An important aim, as with all psychotherapy, is to bring about change. Positive change may occur when a client can direct their strong feelings into making art and when the therapist helps the client to tell their story through the art. How, when, and if change occurs obviously depends on their capacity to engage with this process and needs much time and patience while the client builds confidence. For verbally inarticulate clients, or those who use words defensively, engagement with the art materials gives the opportunity to understand self and environment, communicate emotions to the therapist, receive feedback, and encouragement.

The historical development of art therapy in the United Kingdom, United States, and Europe

There are parallels in the development of art therapy in the United Kingdom and United States, early history being shared with that of group analytic psychotherapy as a phenomenon of the Second World War rehabilitation movement.[5–7] In the 1940s and 1950s art therapists were simply artists working in hospitals who emphasized the healing role of art. In 1963, the British Association of Art Therapists was formed from this small group of artists and art educators, who set themselves the task of defining and extending the activity, preparing standards for training in the higher education sector, informing the public and other professionals of the potential of art therapy, and working towards a career and salary structure in the National Health Service (NHS). The first *postgraduate trainings* began in the late 1960s. The positive response of the NHS and other organizations to art therapy's beneficial impact led to a petition being made for statutory regulation under the old Council for Professions Supplementary to Medicine in 1991, approved in 1997, after which art therapists, along with music and dramatherapists had their own federal Board at the Council. They were transferred to the Health Professions Council in 2001. Training in the United Kingdom is now at Master's level, in four universities in England, one in Scotland, one in Northern Ireland, and usually follows a degree in art and design. Study of psychotherapeutic principles, visual art, and practical placement are important elements in the training. Elsewhere in Europe the picture is very different with some countries sharing the UK standards, others having no training or a great variety of trainings in both the public and private sector. The United States, Australia, and New Zealand have the same requirements of a Master's level qualification in order to practice. (See website references for more information.)

The development of art therapy with specific client groups

One of art therapy's main advantages as a treatment is its flexibility. It can be used with many different client groups and some of these are discussed as follows:

Children

Many founder art therapists in the United Kingdom and United States were art teachers and were influenced by the 'child-centred' approach to art education that developed in the 1930s. American pioneer Kramer considered that it was art activity itself that had inherent healing properties; and that within a secure relationship with the therapist, a child could sublimate their destructive and aggressive feelings by producing an object, which would symbolize those feelings, prevent them being acted out and lead to more insight and control. This often led to change in behaviour.[8]

Others pioneers from the United States gave examples of how group work could enable angry and shameful feelings to be shared among the group members as well as the therapist, to the relief of the child as well as his peers.[9] Many art therapists specializing in work with children attest to the importance of play and to the role of art materials in allowing regression in the form of mess-making. This seems to be particularly beneficial for children who have suffered sexual abuse[10–12] due to the loosening of control that happens when a child becomes deeply immersed in the physical process of painting and is able to lower defences as a result. Materials may be smeared, spilled, and wasted and it is important that the therapist maintains control of the boundaries and is able to tolerate a high level of anxiety as the child attacks the therapeutic space.[13]

Art therapy is helpful for children suffering from chronic constipation, faecal overflow soiling, and also 'antisocial behaviour' and Aldridge[14] pointed out the relationship she observed between food, painting, and faeces while working with neglected and abused children in the context of a social services Unit and how mess-making was important in their creative development. Ambridge[15] discussed how images may be used to reflect mother–child relationships with children who have been sexually abused and are often so traumatized that they cannot speak about their experiences.

The physical involvement in the art materials in enabling regression and essentially in receiving containment and acceptance from the therapist is very important to all the children mentioned above.

Dubowski[16,17] used a Developmental Art Therapy approach in research with children with learning difficulties, aiming to help the child to achieve his or her maximum potential. Understanding creativity and mark-making in early childhood is as important in this model as understanding psychodynamics. Studies made by Kellog[18] of over 100 000 children's scribbles inform our understanding of the developmental process leading to production of meaningful marks. Visual problem-solving through picture-making is developed between the age of about 18 months (when hand-eye co-ordination has developed to the extent that they can grasp an implement and direct it to a picture surface while attending to the activity) and 4 years, by which time most have developed the capacity to make recognizable pictures endowed with symbolic meaning.[19] This model draws on insights from art educationists, most recently Matthews.[20,21] Art therapists have also contributed to the emotional and educational development of children with Autism.[22,23]

Art therapists also occasionally work with families and this is an emerging area of interest.

People with learning difficulties

Stott and Males[24] were among the first British art therapists to write about their work in a large hospital with people who had

lived in institutions most of their lives. They suggested that art therapy offered a means of communication and of self-expression through which difficulties of life in an institution, such as loss of identity, could be eased. Their goals were: to find the art medium of most use to each resident, bearing in mind any physical handicaps; to set-up the art therapy sessions at regular times and in the same place; record and report the results of sessions; to enable maximum communication to take place. For some long-term residents, the art therapy studio became an 'oasis' in the desert of the hospital and they gained an identity as 'artist'.

Their work was continued by a generation of art therapists concerned about the effects of institutionalization on long-stay residents. Drawing on Gardner's work concerning the categories of personal and spatial intelligence[25] Rees set-up a qualitative research project using a detailed observational schedule with a group of women with severe learning difficulties in a single-sex locked ward, conducted over 3 years to investigate clients' use of physical space and its potential symbolic significance. Rees found that by relating to physical and spatial aspects of their environment, some clients discovered an effective way of maintaining some level of psychological and emotional integration. They were helped to manage their often overwhelming feelings and to develop a stronger identity in their, albeit, very restricted environment.[26] Strand used a group interactive art therapy model with a group of learning disabled clients whom she observed to be suffering from loss and despair as a result of their emotional needs being neglected.[27]

Now that people with learning difficulties mainly live in the community, art therapists are now able to assist clients in developing resources to manage their day-to-day activities, and to improve their quality of living—particularly social interaction—through engagement in creative activity, often in groups. Insights from earlier work on institutionalization and exclusion now inform art therapy with older residents in care homes.

Offenders

Liebmann used art therapy within probation services to address 'offending behaviour' directly[28] summarizing the benefits to offenders as follows: as a means of non-verbal communication, important for the high percentage of offenders who have poor verbal skills, conversely with those who use words defensively; to release angry and aggressive as well as shameful and embarrassed feelings and provide an acceptable way of looking at and dealing with difficult emotions; client and therapist together could look back at the images over a series of sessions, see patterns and note developments; active participation is required, helping to mobilize those who may not be voluntary clients and bring about favourable behavioural changes that outlast the session itself. Liebmann devised strategies to help offenders gain insight into their behaviour, for example, the comic strip where the client is asked to draw an important life event within frames, the sequences of which can then be discussed and alternative options suggested.[29] This approach combines some elements of cognitive behavioural therapy with a psychodynamic approach, providing a structure, which is empathetic but on the other hand does not collude with offending behaviour, nor accept it as inevitable.

Teasdale produced a set of Guidelines while working in the prison service. These focus on inmates and on the prison environment, reinforce many of the points above, such as using art therapy to support prisoners in coping with their imprisonment, address feelings of separation, isolation, loss, low self-esteem. Specific advice is offered for conducting art therapy within a prison context.[30]

People with psychotic illness

From the 1940s art therapists have worked with long-stay clients in psychiatric hospitals modifying the effects of institutionalization through detailed attention to communication when even the most psychotic patients could communicate through images and have this acknowledged. Now that people with psychosis increasingly live in the community and cope with the challenges of everyday life the focus is on issues of independence, isolation, building relationships, managing the illness, and its impact on self and family.[31]

Work with acutely psychotic clients, on the other hand, takes place on wards, is normally brief and aimed at helping the client to interact positively with others usually in 'open' studio groups with a rapidly changing population, or for those in a serious acute state, on a one-to-one basis. Drawing on process-oriented psychology McClelland[32] has devised a new model using art therapy to work directly on the acute state itself, requiring an active and assertive therapist style and meeting the client in their own 'language' however bizarre this may seem.

Older people

With older people's mental health and well-being coming under increased government scrutiny, this is an area where art therapy has much potential to be beneficial. Art therapists in this field have noted that attention to loss, fear of illness and dying, feelings of helplessness and dependency is necessary in alleviating depression. Recent art therapy research with older people with dementia showed some improvements in mental acuity, calmness, sociability, and physical competence following 40 weekly sessions of group work, compared to no change in the control.[33]

As with other client groups, engaging in creative activity can also provide an outlet for frustrations to do with ageing, as well as possibly leading to a challenging and rewarding hobby.

Physical illness

Art therapy is used with people suffering from cancer, including terminal cancer[34,35] as well as with other long-term and progressive illnesses or conditions such as multiple sclerosis, ME, kidney disease, effects of stroke, Parkinson's disease, and in palliative care to manage anxiety and fear about death and dying. The aims are to assist clients in coping with the emotional impact of their illness, enabling expression of feeling, relieving depression and stress, and improving their quality of living.

Other areas where promising work is going on: with eating disorders, drug and alcohol addiction, and with refugees and asylum seekers, many of whom have been the victims of war, of torture and are suffering extreme stress; also with the moderately to severely depressed and those who have work and relationship difficulties.

Contextual issues

In the United Kingdom, over 50 per cent of art therapists work within the NHS where they may form part of the psychological therapies or occupational therapy department, or be autonomous. Others practise in Social Services, Education, Home Office,

non-Statutory services, and a fairly small percentage work privately or are self-employed.

All are capable of assessing the suitability of the client for art therapy. Exclusion criteria are minimal. Initial interviews usually explain that the client does not have to be 'good at art' and that art therapy is not a 'painting class' but that it may arouse strong and sometimes difficult feelings. All therapists are trained to work with individuals and groups, to function as members of multi-disciplinary teams, to manage health and safety issues concerning preparation and maintenance of the art therapy space, liaising regularly with medical, nursing, teaching, or other appropriate staff in the interests of the client. Ethical standards are laid down by the Health Professions Council as for the other arts therapists (drama and music and also dance) with whom there is regular contact.

Research

There is a substantial body of qualitative research in art therapy, however much of this would not meet the criteria for evidence-based practice required by today's public sector. Currently there is very little about art therapy within National Institute of Clinical Excellence guidelines, which have tended to emphasize quantitative studies. This is unfortunate as it gives the impression of a profession without a strong evidence base, which is not so as the flourishing Art Therapy Practice Research Network and a growing body of literature demonstrates.[36,37]

A few books and papers emerging from control group studies feature evaluations of art therapy groups with older people with moderate to severe dementia[38] and with schizophrenia. In 2006, the UK Health Technology Assessment supported a consortium headed by Crawford, Killaspy, and Waller (University of London) for a multi-centre random control group study of art therapy and schizophrenia. The Master's level training in art therapy requires a substantial dissertation, and a significant percentage of art therapists continue to doctorate research designed to use and test the hypotheses emerging from over 60 years of detailed casework.

Further information

British Association of Art Therapists: www.baat.org.

Health Professions Council: www.hpc-uk.org (Art, Drama and Music Therapy).

International Society for the Study of the Psychopathology of Expression and Art Therapy: http//www.online-art-therapy.com/

American Association of Art Therapists: www.arttherapy.org.

References

1. Quality Assurance Agency. (2004). *Benchmark statements for arts therapies*, www.qaa.ac.uk.
2. Dalley, T. (ed.) (1987). *Art as therapy*, pp. 6–19. Tavistock, London.
3. Schaverien, J. (1987). The scapegoat and the talisman: transference in art therapy. In *Images of art therapy* (eds. T. Dalley, *et al.*), pp. 74–108. Tavistock, London.
4. Schaverien, J. (1992). *The revealing image: analytical art therapy in theory and practice*. Routledge, London.
5. Waller, D. (1991). *Becoming a profession: the history of art therapy in Britain*. Routledge, London.
6. Waller, D. (2004). *Art therapists: pragmatic rebels*. Goldsmiths College, London (Inaugural Lecture May 2001).
7. Hogan, S. (2001). *The healing arts: the history of art therapy*. Jessica Kingsley, London.
8. Kramer, E. (1971). *Art therapy with children*. Schocken Books, New York.
9. Rubin, J. (1978). *Child art therapy*. Van Nostrand Reinhold, New York.
10. Lee Drucker, K. (2001). Why can't she control herself? Case study. In *Art therapy with young survivors of sexual abuse* (ed. J. Murphy), pp. 101–25. Brunner-Routledge, Hove and New York.
11. Lillitos, A. (1990). Control, uncontrol, order and chaos: working with children with intestinal motility problems. In *Working with children in art therapy* (ed. C. Case), pp. 72–88. Routledge, London.
12. Sagar, C. (1990). Working with cases of child sexual abuse. In *Working with children in art therapy* (ed. C. Case), pp. 89–114. Routledge, London.
13. Waller, D. (2006). Art therapy and Children: how it leads to change. *Clinical Child Psychology and Psychiatry*, **11**, 271–82.
14. Aldridge, F. (1998). Chocolate or shit: aesthetics and cultural poverty in art therapy with children. *Inscape*, **3**, 2–9.
15. Ambridge, M. (2001). Using the reflective image within the mother-child relationship. In *Art therapy with young survivors of sexual abuse* (ed. J. Murphy), pp. 69–85. Brunner-Routledge, Hove.
16. Dubowski, J. (1984). Alternative models for describing the development from scribble to representation in children's graphic work. In *Art as therapy* (ed. T. Dalley), pp. 45–61. Tavistock, London.
17. Dubowski, J. (1989). Art versus language (separate development during childhood). In *Working with children in art therapy* (ed. C. Case), pp. 7–22. Tavistock/Routledge, London.
18. Kellog, R. (1970). *Analysing children's art*. National Press Books, California.
19. Dubowski, J. and James, J. (1998). Arts therapies with children with learning difficulties. In *Development and diversity: new applications in art therapy* (ed. D. Sandle), pp. 41–56. Free Association Books, London and New York.
20. Matthews, J. (1999). *The art of childhood and adolescence: the construction of meaning*. Falmer Press, London.
21. Matthews, J. (2003). *Drawing and painting: children and visual representation*. Paul Chapman, London.
22. Evans, K. and Rutten-Saris, M. (1998). Shaping vitality affects: enriching communication: art therapy for children with autism. In *Development and diversity: new applications in art therapy* (ed. D. Sandle), pp. 57–77. Free Association Books, London and New York.
23. Tipple, R. (2003). The interpretation of children's art work in a paediatric disability setting. *Inscape*, **8**, 48–59.
24. Stott, J. and Males, B. (1984). Art therapy for people who are mentally handicapped. In *Art as therapy* (ed. T. Dalley), pp. 111–26. Tavistock, London.
25. Gardner, H. (1984). *Frames of mind: the theory of multiple intelligences*. Paladin, London.
26. Rees, M. (1995). Making sense of marking space: researching art therapy with people who have severe learning difficulties. In *Art and music therapy and research* (eds. A. Gilroy and C. Lee), pp. 117–37. Routledge, London and New York.
27. Strand, S. (1990). Counteracting isolation: group art therapy for people with learning difficulties. *Group Analysis*, **23**, 255–63.
28. Liebmann, M. (ed.) (1994). *Art therapy with offenders*. Jessica Kingsley, London.
29. Liebmann, M. (1998). Art therapy with offenders on probation. In *Development and diversity: new applications in art therapy* (ed. D. Sandle), pp. 104–20. Free Association Book, London.
30. Teasdale, C. (2002). *Guidelines for arts therapists working in prisons*. Department for Education and Skills/HM Prison Service, Prisoners' Learning and Skills Unit.
31. Killick, K. and Schaverien, J. (1997). *Art psychotherapy and psychosis*. Routledge, London and New York.

32. McClelland, S. (1992). Brief art therapy in acute stages: a process-oriented approach. In *Art therapy: a handbook* (eds. D. Waller and A. Gilroy), pp. 189–208. Open University Press, Buckingham.
33. Rusted, J., Sheppard, L., and Waller, D. (2006). A multi-centre randomized control group trial on the use of art therapy for older people with dementia. *Group Analysis*, **39**, 517–36.
34. Pratt, M. and Wood, M. (eds.) (1998). *Art therapy in palliative care*. Routledge, London and New York.
35. Waller, D. and Sibbett, C. (eds.) (2005). *Art therapy and cancer care*. McGraw-Hill, Maidenhead.
36. Karkou, V. and Sanderson, P. (2006). *Arts therapies: a research based map of the field*. Elsever, London.
37. Gilroy, A. (2006). *Art therapy, research and evidence based practice*. Sage, London.
38. Waller, D. (ed.) (2002). *Arts therapies and progressive illness*. Brunner-Routledge, Hove and New York.

6.5

Indigenous, folk healing practices

Wen-Shing Tseng

What are indigenous, folk healing practices?

Indigenous, folk healing practices are nonorthodox therapeutic practices based on indigenous cultural traditions, operating outside of official (modern) healthcare systems.[1] These practices are often validated by experience, but are not founded on scientific principles. Indigenous healing practices are observed in 'primitive' or 'pre-industrialized' societies as well as in modern or developed societies. All healing practices or psychotherapies are more or less culturally influenced, including modern and orthodox psychotherapies, but indigenous healing practices are described as 'culturally embedded' because they are often intensely embedded in the cultural systems in which they were invented and in which they are practised. They are, therefore, usually very difficult to transplant to entirely different cultural settings, where they do not have the same meaning or legitimacy.[2]

While indigenous healing practices function in general as healing methods for problems, they are not usually considered by either the healer or the clients to be psychological therapy for the clients' emotional or psychological problems. Rather, they are recognized as religious ceremonies or healing exercises related to supernatural or natural powers. However, from a mental health point of view, the indigenous healing practices often provide psychotherapeutic effects for the clients, and can be considered as folk psychotherapy.

Anthropologists have studied folk healing practices as a part of cultural behaviour. Recently, cultural psychiatrists have become interested in examining indigenous healing practices from clinical perspectives to explore the similarities and differences that exist between folk healing practices and modern psychotherapy, and to disclose the therapeutic mechanisms that are operating in and being utilized by indigenous healing practices. Many people in developed societies utilize folk healing practices as adjunctive to their primary (modern) therapy or as their main way to get help. Therefore, it is relevant for the modern psychiatrists to know what they are and the possible therapeutic mechanism they offer, or the possible negative effects they may receive by utilizing such indigenous healing practices.

Various practices are covered by the loosely defined terms, indigenous or folk healing. Religious healing practices and ceremonies are closely related to a specific religion. Shamanism involves a spirit medium. Divination, or various kinds of fortune-telling, including astrology or physiognomy, may be used by people to solve their psychological problems or to seek answers for life problems, and, therefore, can be viewed as folk counselling practices as well. Furthermore, the practice of meditation, a self-training exercise used to obtain tranquility, growth of mind, and prevention of emotional problems, can be considered a folk healing practice if one defines psychotherapy very broadly, as not only treating a suffering person but also providing a means for preventing problems and improving the quality of a person's mental life.[3]

No matter what terms are used, indigenous healing practices share some common features. They are invented and utilized by local people for the purpose of solving problems or treating suffering—therefore, they are called indigenous in contrast to universal. They are distinctly different from the modern (Western or orthodox) professional medical approaches—thus, they are called folk practices. Most of them are supernaturally oriented and remote from any scientific orientation. Such indigenous practices are usually rooted in traditional beliefs and folk interpretations of problems, and, thus, are closely related to cultural beliefs.

Subdivision of various healing practices

Based on their core nature and their basic therapeutic orientation, healing practices observed in different societies can be subdivided into different categories namely, supernatural orientation (such as spirit mediumship, religious healing ceremony, and divination); nature orientation (such as fortune-telling, astrology, and meditation); medical-physiological orientation (such as mesmerism, acupuncture, and herb medicine); and socio-psychological orientation (such as Zen training, Alcoholics Anonymous, est, and most modern psychotherapy).[4,5] It is recognized that such subdivisions are arbitrary, and often overlap. Yet these subcategories will help us to understand various healing practices that exist on a spectrum which includes the supernatural, natural, physiological, and psychological.

Spirit mediumship (trance-based healing system)

Spirit mediumship broadly refers to a situation in which the healer or the client, or both, experiences alternate states of consciousness in the form of dissociation or a possessed state at the time of the healing ritual. From a psychotherapeutic point of view, it is

important to distinguish which person is in an alternative state of consciousness, as the mechanism of therapy differs depending on whether it is the healer or the client who is dissociating.

(a) Shamanism

It is speculated that the geographic heartland of shamanism is Central and North Eurasia, with widespread diffusion to Southeast Asia and the Americas.[3] Through a religious ceremony, a shaman can work himself into a trance state in which he is possessed by a god. The rhythmic singing, dancing, or praying (quiet meditation) seems to assist the self-induction of the trance state. Among native healers in North and South America, a psychedelic substance (such as may be found in cactus) is frequently used to induce an altered state of consciousness and a special psychic experience for the healing performance. Whether the altered state of consciousness is substance- or self-induced, the healer is considered to be possessed by a supernatural power. The client can then consult the supernatural through the shaman for instructions on dealing with his or her problems.

The causes of problems are usually interpreted according to the folk concepts held by the culture—involving such things as loss of the soul, sorcery, spirit intrusion, or violation of taboos. Disharmony with nature may also be interpreted as the cause of problems. Coping methods are usually magical in nature, such as: prayer, the use of charms, or the performance of a ritual ceremony for extraction or exorcism. Utilizing supernatural powers, acting as an authority figure, making suggestions, and providing hope are some of the main mechanisms for healing provided by the shaman. The goal of the healing practice is to resolve the problems that a client is encountering.

(b) Zar ceremonies

The term *zar* refers to a ceremony as well as a class of spirits. *Zar* ritual is observed primarily in Muslim societies in the Mideast, including Ethiopia, Egypt, Iraq, Kuwait, Sudan, and Somaliland. *Zar* ritual is different from shamanism in that, in addition to the healer, the client also experiences the dissociated or possessed state.

The *zar* ceremony is primarily a female activity. All of those attending the ceremony wear new or clean clothing to please the spirits. The main patient usually wears a white gown, as much gold jewelry as possible, and is heavily perfumed. The ceremony master begins the ceremony with a song and drumming. When a spirit associated with some person in the audience is called, that person begins to shake in her seat, dancing, and trembling until she falls, exhausted, to the floor. Before the spirit consents to leave, it usually demands special favours, such as jewellery, new clothing, or expensive foods. It is the duty of the relatives and friends to gather around the prostrate woman and pacify the spirit. The whole tone of the ceremony is one of propitiation and persuasion, rather than coercion. The ceremony ends with an animal sacrifice and a feast.

The *zar* ceremony is primarily an adult female activity reflecting social conditions of sex-separation, low female status, restriction of women from religious participation, an unbalanced sex ratio, marital insecurity, and relative isolation. The *zar* ceremony provides women an ideal situation for relief of persistent and regular anxieties and tensions arising from their life conditions. The goods demanded during the ceremony are all things that their husbands should provide. This fulfills a woman's wish for attention and care. Emotional catharsis, fulfillment of unsatisfied desire, and compensation for the suppressed female role are some of the therapeutic mechanisms working in this kind of therapeutic ritual. Restoring balance in real life is the implicit goal of this culture-embedded healing practice.[6]

Religious healing ceremonies

A distinction needs to be made between religion and a religious healing ceremony. Religion refers to a system of belief in a divine or superhuman power or spiritual practice. As a part of a religion, some people may perform special ceremonies for the purpose of healing certain problems or disorders. There are various kinds of religious healing ceremonies observed in different societies that are considered by mental health workers to serve a therapeutic function for their participants.

(a) Sprit dancing ceremony

As observed among Salish-speaking Indians of the Pacific Coast of North America, healing ceremonies utilize psychological mechanisms and processes similar to those in brainwashing. The initiate has to go through three major therapeutic approaches: depatterning through shock treatment (such as physical restraint, blindfolding, hitting, kinetic stimulation, or intensive acoustic stimulation, followed by lying still, being forbidden to talk, and starvation); physical training (such as daily running, jumping into ice-cold waters, or frequent rounds of dancing); and, finally, indoctrination.[7]

(b) Sacrificial ritual

An example is found in Yoruba, Africa. A person's problems were identified by palm nuts tossed by the diviner. It was usually interpreted that the person or a member of his family had offended the family *orisa* (the lineage deity) or some other spirit. Then, the sacrifice of a certain animal was prescribed for resolution. In the sacrifice, the supplicant passed his bad luck or illness to the animal, and the animal was killed in the supplicant's stead. The healing power of the ritual lies in its reassurance and generation of conviction. The ritual demonstrates that proper curative steps are being taken.[8]

(c) The religious ceremony of mourning

This is practised among members of the Spiritual Baptist Church in the West Indies. In a desire for spiritual strength and other benefits, church members volunteer to participate in the practice. After ceremonial washing and anointing, the mourners are isolated in a small chamber at the back of the church, where they remain for a period of 7 days. During that time, each individual prays, fasts, and experiences dreams and visions. The mourners claim they obtained beneficial psychological relief on their moods; attainment of the ability to foresee and avoid danger; improvement in their decision-making abilities; cures for physical illnesses; and heightened facility to communicate with God.[9]

(d) Snake-handling cult

An extremely different form of religious ritual was a snake-handling cult in the southern United States. As part of cult activities, members, in trance states, handled poisonous snakes as a sign of being blessed by God. Occasionally, some of the members died when they were bitten by the snakes. Although forbidden by the government to perform such cult rituals, these activities still exists. The gratification of emotional excitement was interpreted

as one of the effects sought by cult members—even at the risk of their lives.[10]

(e) Christian religious healing

It is important to know that religious healing ceremonies are not only observed in primitive societies or among uncivilized populations, but are quite common in many industrialized societies, as well. In Christian religious healing, there is a broad spectrum of beliefs and activities, ranging from Christian Science to the fundamentalism of healers such as Oral Roberts to the Roman Catholic rite of anointing of the sick. The participating client is provided with the hope for supernatural resources against disease, thus increasing his or her security and sense of well-being.[11]

Thus, in various forms of religious healing ceremonies, the therapeutic operation is carried out through the ritual of prayer, testimony, sacrifice, reliving experience, or even spirit possession. Assurance, suggestions, and generation of conviction are some of the healing mechanisms utilized in the practices. The aims of therapy are to heal the problems and give a certain perspective to the client's life.

Divination

Divination refers to the act or practice of trying to foretell the future or the unknown by occult means. It relies on mysterious, magic, or religious methods. Since the interpretation of divine instruction is usually provided by the diviner himself, or an interpreter, the interaction between the diviner/interpreter and the client becomes an important variable.

There is a range of methods of divination. Some methods are very simple, while others are more complicated. For example, in Nigeria, Africa, the divination practised by the Nsukka Ibo (called *Afa*) is carried out by casting four strings containing half-shells of the seeds of the bush mango;[12] and by the people in Yoruba (known as *Ifa*) by tossing palm nuts.[8] In the divination practised in some parts of Africa, the diviner simply offers a certain sign himself. For example, divination may occur while his hand is shaking, with the belief that he is guided by a supernatural power to give instructions.

In ancient China, turtle shells or the bones of big animals were burned during divination ceremonies and divine instruction was interpreted through the cracks made from the heat. An elaborate divination system called *chien* has been developed in China, and a modified version is used in Japan. To obtain answers to questions about their lives, some Chinese or Japanese will visit temples for divination. After a sincere prayer to the god of the temple, the person will ask for divine instruction, which is provided through a fortune stick that the person selects. Corresponding to the number on the stick, there is a fortune paper with an answer written on it. This practice is called *chien* drawing in Chinese,[13] or *kujibiki* in Japanese.

No matter what method of divination is practised , the basic therapeutic operation is performed to provide a clear-cut answer for the problems presented. Thus, it is helpful psychologically for a client to find a definite way to address his problems. Naming effects are among the healing mechanisms operating in divination. It is assumed that human life is under the influence of supernatural regulation. It is the basic goal of the person seeking help to find the proper way to comply with the universe through divine instruction.

(a) Fortune-telling

The system of reference shifts from the supernatural to the natural in the practice of fortune-telling. Based on the concepts of microcosm and macrocosm, fortune-telling is oriented to the basic belief that human life and behaviour are parts of the universe. The nature of the problems is usually explained in terms of an imbalance of vital forces or disharmony with the natural principles that rule the universe. The objective of the practice is to help the client find out how to live compatibly with nature and adjust to the environment more harmoniously.

Based on the sources of information used, fortune-telling can be divided into several groups. In astrology, there is a basic belief that a person's life is correlated to and influenced by the movement of the stars, thus, their movement becomes the essential source of information for predicting one's life course. For the Chinese, an ancient record of universal change, the Oracle of Change (*Yi-Jing*), is used for fortune-telling. A person's date and time of birth, and the number of strokes in the Chinese character for his or her name, is the information needed to calculate an individual's fortune.

Physiognomy is based on the assumption that there is a close correlation between the mind and the body and that one's character, life, and fortune can be read by examining one's physical features. It is assumed that a person is born with a certain predisposition, which is shown in his physical appearance and will lead him to manifest certain behaviour patterns. A physiognomist tries to help a client understand his own character and behaviour patterns, learning how to make good use of his talents and, at the same time, make up for his shortcomings.

Although the basic assumption underlying fortune-telling is that every person has a predetermined course of life, such fate is not absolutely unchangeable—it may be subject to modification. Thus, it is not a completely passive acceptance of fate, but allows room for adjustment. Finding a way to adjust your own fortune is the purpose of fortune-telling.

Even though the basic orientation shifts from a supernatural to a natural one, and the sources of practice rely on the rules of nature, the therapeutic operation, like divination, is still characterized by offering folk-natured interpretation and providing concrete guidance for a client in making choices. Based on the concepts of microcosm and macrocosm, complying with the fundamental rules of nature is the basic goal of the practice.

Common therapeutic factors

Reviewing various forms of folk therapy, it has been pointed out that the core of the effectiveness of different methods of religious and magical healing seems to lie in their ability to arouse hope by capitalizing on the patient's dependency on others.[14] Comparing the healing practices carried out by witch doctors and psychiatrists, it has been pointed out that they share a common root. Both kinds of therapists are able to decrease the client's anxiety by identifying what is wrong with him—that is, to name the cause of the problems, providing the effect of the Rumpelstiltskin principle; the therapist presents certain personal qualities that are admired by the culture and contribute to the therapy; the client's expectations of therapy, and the emotional arousal that is usually enhanced by the therapeutic setting, the therapist's belief in himself, and his reputation; the emerging sense of learning and mastery that the

client obtains through therapy; and finally, the techniques of therapy that enhance the basic components of psychotherapy. It is clear that folk healing practices and modern psychotherapy share a number of nonspecific therapeutic mechanisms.[15] It has been indicated that traditional healing practices have several advantages over cosmopolitan modern medicine, namely: cultural congeniality, maximal use of the personality of the healer, a holistic approach, accessibility and availability (particularly for developing areas), effective use of affect and altered states of consciousness, collective therapy management, and cost-effectiveness.[1] It is important to recognize that both folk healing and modern therapy utilize symbols and metaphors for interpretation and suggestions.[16] In contrast to modern psychotherapists, some folk healers make use of symbolic interpretations and suggestions to enhance the effects of healing.

Attitudes towards indigenous healing practices

Different points of view exist among scholars, clinicians, and public health workers regarding whether or not to encourage or discourage indigenous folk healing practices in various societies. Some people (particularly modern clinicians) see folk healing as merely superstitious and primitive, insisting that such out-of-date practices should be discouraged or prohibited. Others (such as cultural anthropologists and cultural psychiatrists) consider these folk practices to be interesting subjects for academic study—examining the therapeutic elements that are utilized in these primitive healing practices, and why such supernaturally oriented therapeutic exercises are still popular among some groups. Still other people (such as some community health workers) believe that, due to the shortage of professional personnel available in the community, the existence of folk therapies should be supported. The position was taken that any folk healing practice that is proven (or at least considered) to be helpful to the client and useful to the community deserves the support and encouragement of clinicians as well as administrators.

Final comments: clinical implications

The comparative study of indigenous healing practices and modern psychotherapy has revealed the existence of certain universal elements of the healing process that operate as important factors for therapy, whether the therapy is carried out in a primitive or modern form. The universal and nonspecific healing factors identified are: the cultivation of hope, the activation of surrounding support, and the enhancement of culturally sanctioned coping. The study of indigenous healing practices has also pointed out the existence of supernatural dimensions of healing power, which are less intentionally utilized in modern therapy.

Despite the general usefulness of folk therapies, the ill effects of some have not been widely studied and reported. Yet, clinical observation has disclosed that some folk therapists cause harm to the clients who seek their services. Under the guise of treatment, tricking a client out of his money by deceit or fraud, or sexual involvement with a client, are examples of disreputable behaviour that are occasionally reported. Harming a client by prescribing dangerous substances, and physically injuring or even killing a client by accident during the performance of an exorcism, are other examples of serious complications that have occurred.

No matter what position is taken, there is one simple fact that deserves attention, namely, that there exists a wide range of professional quality among so-called folk healers, and different motivations for practice. Some are benign healers motivated by a desire to serve, while others are not. Some are well-trained in their particular professions and know how to practice within its limitations, while others are not—and are liable for malpractice. The major problem is that, from a public health point of view, in most societies, there still are no formal guidelines for regulating folk therapy, as there are for modern therapy. Folk therapy, whether it is shamanistic practice or faith healing, should be subject to periodic surveys and reevaluation by the public health administration, as is modern clinical work, so that its benefits to clients can be protected and any potential malpractice can be prevented. If any folk therapist refuses to be examined and regulated, he or she should be discouraged or prevented from practicing.

Further information

Tseng, W.S. (1999). Culture and psychotherapy: review and practical guidance. *Transcultural Psychiatry*, **36**(2), 131–79.

Tseng, W.S. (2003). 7: Culturally competent psychotherapy. In *Clinician's guide to cultural psychiatry* (ed. W.S. Tseng), pp. 291–342. Academic Press, San Diego.

References

1. Jilek, W.G. (1994). Traditional healing in the prevention and treatment of alcohol and drug abuse. *Transcultural Psychiatric Research Review*, **31**(3), 219–58.
2. Tseng, W.S. (2001). *Handbook of cultural psychiatry*, pp. 515–37. Academic Press, San Diego.
3. Prince, R. (1980). Variations in psychotherapy procedures. In *Handbook of cross-cultural psychology: psychopathology*, Vol. 6 (eds. H.C. Triandis and J.G. Draguns). Allyn and Bacon, Boston.
4. Tseng, W.S. and Hsu, J. (1979). Culture and psychotherapy. In *Perspectives on cross-cultural psychology* (eds. A.J. Marsella, R.G. Tharp, and T.J. Ciborowski). Academic Press, New York.
5. Tseng, W.S. (1999). Culture and psychotherapy: review and practical guidance. *Transcultural Psychiatry*, **36**(2), 131–79.
6. Kennedy, J.G. (1967). Nubian *zar* ceremonies as psychotherapy. *Human Organization*, **26**(4), 185–94.
7. Jilek, W.G. (1976). Brainwashing as therapeutic technique in contemporary Canadian Indian sprit dancing: a case in theory building. In *Anthropology and mental health: setting a new course* (ed. J. Westermeyer). Mounton Publishers, Paris.
8. Prince, R. (1975). Symbols and psychotherapy: the examples of Yoruba sacrificial ritual. *Journal of American Academy of Psychoanalysis*, **3**(3), 321–38.
9. Griffith, E.E.H. and Mahy, G.E. (1984). Psychological benefits of spiritual Baptist mourning. *The American Journal of Psychiatry*, **141**(6), 769–73.
10. La Barre, E.H. (1962). *They shall take up serpents: psychology of the southern snake-handling cult*. University of Minnesota Press, Minneapolis.
11. Hufford, D. (1977). Christian religious healing. *Journal of Operational Psychiatry*, **8**(2), 22–7.
12. Shelton, A.J. (1965). The meaning and method of *Afa* divination among the northern Nsukka Ibo. *American Anthropologist*, **67**, 1441–5.

13. Hsu, J. (1976). Counseling in the Chinese temple: a psychological study of divination by Chien drawing. In *Culture-bound syndromes, ethnopsychiatry, and alternate therapies* (ed. W.P. Lebra). University Press of Hawaii, Honolulu.
14. Frank, J.D. (1961). *Persuasion and healing: a comparative study of psychotherapy*. Schocken Books, New York.
15. Torrey, E.F. (1986). *Witchdoctors and psychiatrists: the common roots of psychotherapy and its future*. Harper & Row Publishers, New York.
16. Kirmayer, L.J. (1993). Healing and the invention of metaphor: the effectiveness of symbols revisited. *Culture, Medicine and Psychiatry*, **17**(2), 161–95.

SECTION 7

Social Psychiatry and Service Provision

This page intentionally left blank

7.1

Public policy and mental health

Matt Muijen and Andrew McCulloch

Introduction

Public policy, and specifically national public policy, is one of the key factors that affects the practice of psychiatry, the shape of mental health services, and the environment within which mental health services work. The specific content of public policy varies greatly across the world and often even across neighbouring countries. It is therefore impossible within the space of this chapter to undertake systematic international comparisons. This chapter gives an overview of:

(a) what policy is and why it might be important;

(b) types of policy and policy development internationally;

(c) international structures and organizations that are relevant to the scope and content of policy, especially in the field of human rights which is often the starting point for policy;

(d) the breadth of policy activity that is relevant to mental health—stretching beyond the health ministry and health policy—and the partnerships that are necessary to tackle the mental health of individuals and populations.

What is public policy?

Koontz and Weihrich[1] define policy or policies as 'General statements or understandings which guide thinking on decision making.' This definition implies that:

1 policy may be stated in writing (statute, guidance, or statement) or may be unwritten and disseminated through management or political chains of command only;

2 policies may exist at various organizational levels albeit here we are mainly concerned with national policies;

3 policy is only one factor in any final management or clinical decision. Other factors will often be more important, and clinical behaviour is notoriously resistant to policy influence except where certain behaviour is directly prohibited.[2]

Therefore, the policy needs to be understood as just one of the factors which determine the nature and state of mental health services and mental health promotion or public mental health within a society.

In many developed countries, the management of the public sector has seen major changes over the last two decades, shifting from a hierarchical and bureaucratic administration to a model of 'managerialism'.[3] In the old model policy making and administration was separated, and governments were responsible for the provision of services often via related public sector agencies. The new model, inspired by business,[4] is characterized by governments introducing competition and market principles and involving the private sector in providing services, setting measurable objectives, and introducing performance management with the aim of delivering higher efficiency. This has affected professionals such as psychiatrists by shifting accountability for good medical practice as judged by peers to reporting on outputs to managers as specified by contracts and practice guidelines. Nevertheless, in most cases strong professional accountability remains, creating a dual accountability that may be appropriate but is always challenging and sometimes conflicting.

Mental health public policy

Public policy in mental health is driven by a range of influences and interests which policy makers will attempt to balance. Positive drivers, that is factors that encourage the development of progressive mental health policies, include:

- epidemiological evidence of the prevalence and the contribution of mental disorders to the burden of disease;
- evidence for effective and efficient interventions;
- public interest as expressed by politicians and the media;
- campaigns by NGOs and professional groups;
- human rights considerations;
- pressure from patient and family organizations;
- growing public expenditure.[5]

The strong interface between mental health issues and human rights explains the importance of human rights conventions and declarations in the shaping of policies and legislation, especially in countries where mental health policy and practice are less developed historically. Negative drivers, that is, factors that limit the commitment of policy makers, include budget limitations,

competition from other priority areas, discrimination, and the perception that mental health is not a fruitful area for investment.

The function of mental health policy

Mental health policy may serve a variety of functions ranging from the purely political and presentational through to delivering on a moral or social imperative. Often policy is mixed in function, enabling politicians to make a statement on an important human rights or social issue whilst also achieving real improvements for people with mental illness and delivering on economic targets. Mental health policy will also affect delivery of other aspects of policy in relation to crime, housing, education, and health. Thus, a coherent set of social policies cannot be developed without addressing mental health in some way. Mental health has been called a 'wicked' social policy issue (see Bogdanor for a discussion of such issues)[6] because it is hard to define, to address, and to deliver on tangibles, but unless we do, many of our social aspirations will not be fully realizable. Finally, the stated mission or desired outcome of policy may vary—for example some countries may have goals in terms of reduced prevalence of an outcome such as suicide, others may wish to achieve de-institutionalization.[7]

Broader public policy and mental health

Whilst policy on mental health services *per se* is likely to be the concern of the health ministry, the impact of other public policy on both mentally ill people, and the practical application of psychiatry may be even bigger. People with severe or persistent common mental health problems often face social exclusion and face difficulties with issues including housing/shelter, employment, education, and welfare or welfare benefits. All mental health professionals in all countries of the world will be aware of how much these issues can affect the lives of people with mental illness and indeed clinical outcomes. Box 7.1.1 summarizes some of the main relevant policy areas which will impact on mental health services.

It should be clear from only brief examination of this table that achieving a totally coherent mental health policy presents huge challenges:

- The views of different parties and the priorities of many different government departments must be reconciled;
- Policy must integrate horizontally—for example, across health, education, and local government, and vertically from small organizations up to government;
- The non-measurable side of what is offered is very important to the stakeholders—services must be responsive, caring, joined up and visible to users, carers, and the general public—yet these features of services are hard to influence from central government;
- The needs and demands are so large and diverse it is difficult to construct and resource a coherent system for delivery even in the richest countries;
- Achieving equity and balancing resource allocation across groups of people suffering from a range of mental disorders with different prevalence and causing a different burden to self, families, and communities. Different groups of disorders require different investment producing different potential benefits yet such decisions must often be made on the basis of ambiguous information whilst subject to conflicting pressures from advocacy groups, communities, media, colleagues, government ministries, and parliament.

Box 7.1.1 Policy links for mental health across government

Government department or ministry (generic description may not accord with structures in all countries)	Mental health relevant policy responsibility (Examples)
Department of health/health and social services	Primary mental health care Human resources for health and social care services Finance for health and social care Specialist mental health care policy Public mental health or mental health promotion
Department of employment/ social welfare	Welfare benefits for people with mental health problems Employment rights and protection Occupational health
Department of education	Education on mental health issues as part of wider curricula in schools and higher education Healthy schools Higher education for relevant vocational qualifications
Department of criminal justice/internal security	Diversion of mentally ill offenders from the criminal justice system Public security
Department of housing/ communities and local government/regions	Housing for mentally ill people Community development Urban policy
Department for constitutional affairs/human rights	Protection of those who lack capacity
Treasury or finance ministry	Financing of the above at a macro level Strategic value for money

(The authors after Jenkins *et al.* (2002))

Comparative policy

It can be argued that mental health policies in most countries fall within four broad groupings:

1 Provision for the basic protection of human rights of people with mental illness or those lacking mental capacity;

2 As one together with a simple health care policy to provide for the health care needs of people with serious mental illnesses;

3 The above but with a more comprehensive health and social care policy which addresses primary and specialist care needs for various groups of disorders;

4 A comprehensive cross-sectoral mental health policy that addresses care, public mental health, and broader issues such as housing.

No country in the world has a fully comprehensive policy although developed countries such as New Zealand, Australia, several

European countries, and states of Canada and the United States have advanced policies. This reflects the nature and resources of these societies, and does not necessarily result in differences in prevalence or outcomes.[8]

Human right conventions and legislation

Human rights have historically been a fundamental driver for the development of mental health policy in most countries of the world. The current Human Rights conventions derive from the United Nations, and are binding once they are ratified by member states following adoption by the United Nations General Assembly. The Universal Declaration of Human Rights was adopted in 1948. This was followed in 1966 by the International Covenant on Political and Civil Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR), in combination known as the International Bill of Rights. None of the conventions directly addresses issues related to mental health care, although the right to the highest attainable standard of physical and mental health is included in Article 12 of the ICESCR. In non-binding general comments, the committee on Economic, Social and Cultural Rights specified that right to health covers availability, accessibility with a strong emphasis on equality and non-discrimination, and acceptability including cultural sensitivity and quality. The obligation of signatories to provide information on detentions and measures taken to prevent abuse for people admitted to mental hospitals is important.[9]

In 1991, the United Nations General Assembly adopted 'the principles for the protection of persons with mental illness and the improvement of mental healthcare', better known as the MI principles, which remains a key reference document for national mental health legislation. The document outlines 25 principles, stressing the importance of non-discriminatory practice within the health care system. The rights to which patients with mental health problems are entitled, according to the MI principles, include the right to live and work as far as possible in the community, and the right to the best available treatment and care in the least restrictive settings in conditions suited to the cultural background of the patient. The principles also set out expectations for confidentiality, the protection of autonomy, and the conditions for involuntary detention. In combination the principles offer a good foundation for modern mental health policies and legislation. Although the document is not legally binding, it is clearly stated that member states are expected to implement these principles fully.

An essential principle, confirmed by the World Conference on Human Rights in 1993, is that all human rights are universal, equally applying to people with disabilities and mental disorders. This is a strong driver for national governments to deliver anti-discrimination legislation, guidance, and practice.

International agencies

There are several relevant regional agencies that represent countries, including political organizations such as the European Union, the Council of Europe and the European Court of Human Rights, the African Commission on Human and People's Rights, and the Inter-American Court of Human Rights. All have produced conventions and declarations pertinent to people with mental health problems, some legally binding.[10]

The World Health Organization (WHO), the specialist health agency of the United Nations, is mandated by its constitution to support member states in all health areas, including mental health. Its core functions include to:

- propose conventions, agreements and regulations, and make recommendations with respect to international health matters;
- assist governments, upon request, in strengthening health services;
- promote improved standards of teaching and training in the health, medical, and related professions;
- study and report on, in cooperation with other specialized agencies where necessary, administrative and social techniques affecting public health and medical care from preventive and curative points of view, including hospital services and social welfare.

In 2001, WHO dedicated the World Health Report to mental health. The report entitled 'New Understanding, New Hope' urges member states 'to seek solutions for mental health that are already available and affordable'.[11] The report formulates a set of 10 recommendations for member states to address the challenges faced in various areas of mental health. By adopting the report (World Health Assembly Resolution WHA 55.10), governments have committed themselves to implement its recommendations to strengthen primary care and develop accessible and affordable community-based mental health services.

At WHO regional level there have been several declarations and action plans that elaborated the principles of the World Health Report 2001, and reinforced the commitment by governments to delivery. Examples include the Caracas Declaration for the Americas, anticipating the World Health Report, and the European Declaration for Mental Health signed in Helsinki 2005.

Why public mental health policy?

Although the governments of most countries accept a role for public policy in the regulation, commissioning, funding, resourcing, and provision of mental health care, particularly for persons with severe and enduring mental health problems, a case still should be made why mental health care cannot be left to a combination of the private market and civil society to deliver without central intervention. As we have argued, when national systems are compared, considerable variation emerges, even within developed countries. In some, the role of government is limited to regulation of independent health insurance groups that purchase from independent providers, and the public funding of a safety net for expensive acute care or long-term conditions. In other countries all parts of the health system are state owned and tax funded.

The World Bank produced a report on this question,[12] scrutinizing the arguments for public financing of mental disorders as compared to other conditions that compete for public money. In their opinion, disease burden, cost-effectiveness of interventions, and externalities are not sufficient arguments since these are not unique to mental health. However, the potential of catastrophic cost, the risk of insurance market failure, and the place of involuntary treatment and consequent issues of human rights are all disproportionately present in mental health care and plead in favour of a public role. One could add to this that stigma and

discrimination may negatively affect the availability and access to mental health services if left to market forces together with the argument advanced above that comprehensive social policies cannot logically leave out mental health.

Economic impact of disease burden

An important argument for the view that mental health should be central to health care and public policy in general is the burden of disease evidence, addressed elsewhere in this book (cross ref). The realization that mental disorders contribute so much to the overall burden of disease puts mental health care at least on an equal footing with other groups of disorders that have long been recognized as posing a major public health challenge such as infectious disease, cancer, and heart disease. Even more startling is the very high proportion of years lived with disability attributable to mental health problems and the days lost to employment due to mental health conditions. The majority of mental health problems contributing to disability are anxiety and depression and the burden of substance misuse and organic disorders in developed countries is also very high, and rising.

From a public policy perspective, this means that the consequences of mental health problems are no longer only an issue of health care, but are central to macro-economic national interests, particularly at times of skills shortages. This has become connected with the interest in the mental well-being of the population, based on the observation that level and growth in Gross National Product (GNP) is only very marginally associated with status of mental well-being of the population, and in some cases even inversely correlated.[13] This has resulted in the conclusion by economists that government's macro-economic policies should not simply be driven by the aim to maximize the Gross National Product and the income of its citizens. Rather, the 'happiness' of the population should become the prime driver of policy making.[14] Governments have consequently seen a need to stimulate the employment of people with mental disorders, both as part of the social inclusion agenda and to reverse the very high and increasing cost of mental disability. This could be considered as a somewhat symptomatic approach ignoring the more systemic problem of job stress and insecurity in very competitive market economies. However, there is also no doubt that loss of employment is a disastrous event for individuals and families in its own right, whether considered from an emotional, social, or financial perspective. The strong association of unemployment with divorce, mental illness, suicide, all factors associated with social exclusion[15] in their own right, supports the case for increasing employment rates, if not exclusively so. A challenge mental health services are facing is to identify the most appropriate and effective response to such broader societal needs as many of the levers for change lie outside the health care sector.

The scope of mental health public policy development

The scope of mental health policy making has broadened, as reflected in the content of international declarations. During the second half of the twentieth century mental health policy and expenditure was focused on people with severe and enduring conditions cared for in institutional settings. Increasingly the focus is broadening to incorporate mental health promotion and prevention and treatment of common mental disorders, including a key role for primary care. The reasons for this include:

- an increased demand for the treatment of stress and depression;
- the development of community-based services, lowering the threshold to access for, and acceptability of care;
- the growing evidence base for a range of interventions such as cognitive behavioural therapy (CBT);
- increased media interest in common mental disorders and their treatment;
- the reduction of stigma of common mental disorders;
- growing affluence in many countries making privately purchased therapies more affordable;
- an awareness that mental health problems are associated with many social determinants of health, including inequality, and macro-economic conditions;
- the co-morbidity between mental health problems and many physical diseases including diabetes and CHD, affecting mortality, and recovery rates;
- an awareness of the cost of mental illness to society, and the evidence of cost-effective promotion and prevention activities.

All these have major consequences for public policy. The complex interface between public mental health activities and mental illness services demands a multi-agency approach, crossing the responsibilities of government departments and local services and requires 'joined up government'. This requires clear policies with explicit objectives and targets, specifying responsibilities for delivery and funding. It requires the active ownership of mental health issues beyond the health care sector and its advocacy by leading practitioners across a range of sectors and disciplines.

Policy implementation

The growing range and complexity of mental health-related activities requires explicit strategies outlining the vision and values of reform, the planned service organization, capital and human resource implications, financing, quality monitoring, and human rights protection. The number of countries in the world with such detailed mental health policies as defined above is still low but growing. There are also many examples of good strategies, whether at national, state, regional, or local level, that have produced impressive change, and guidance for the drafting of policy and legislation is available.[16]

The common factors in countries with a successful track record of change include:

- full commitment by government;
- national consultation exercises, gaining the views and support of key stakeholders;
- clear and consistent communication of objectives;
- realistic and sustained resource allocation;
- workforce strategies including education and training.

However, there are also examples of countries that have drafted comprehensive strategies that have not been implemented.

Failure of implementation will create cynicism that may obviate future attempts. Main factors that lead to failure include:

1 Poor connection between the strategy and reality, and lack of focus. In some instances strategies are written by foreign experts ignorant of local circumstances. Other examples are of strategies that are based on ideology and are overambitious, repeating the content of declarations without taking into account local needs and resources.

2 Lack of financial commitment or poor grasp of the financial costs of a new model of care, causing rejection by the finance ministry or inability to deliver locally.

3 Neglect of the human resource implications. New models tend to require more staff with new skills, and implementation can create high levels of anxiety that need to be confronted.

4 Lack of consultation with professional groups, users, carers, and communities, who therefore reject or resist the strategy.

5 Insufficient joint planning either across government or with partner organizations that will carry responsibilities for the funding and delivery of essential components such as social care, housing, and employment. Incentives will need to be created for various groups such as professionals and the private sector to support the strategy. Added to this is the need to align incentives across different areas, and to prevent the development of perverse incentives, leading to conflicts of interest.[17]

6 Lack of a long-term perspective. It is often possible to develop local model services, but generalization to national and sustainable programmes requires different approaches.

7 Lack of all party political support. It is unlikely reform will be initiated and embedded within the lifetime of a single parliament/administration or under the governance of the same minister of health. Unless political support is strong across the political spectrum, the risk of reversal due to weak commitment or even hostility to previously agreed plans is high.

Human resources

Any strategy must include plans for sufficient staff with the right attitudes, skills, and competencies. The growing demand for mental health care has increased the need for staff. Many countries around the world are looking for ways to remedy staff shortages, particularly of doctors and nurses. Three approaches can be distinguished:

- increase of national recruitment;
- creation of different roles and responsibilities for staff;
- international recruitment.

Although the increase of training places is on the face of it an obvious approach, it does not solve the problem immediately since the lag time between the creation, for example, of a medical training place and the production of a psychiatrist will be at the least 10 years. The availability of training places assumes that places will be filled with able candidates, which is also not always the case. In some poorer countries students reject the option to become mental health workers due to stigma, poor working conditions, and poor pay.

An important alternative strategy is to take a more comprehensive competency-based approach, and analyse the roles and responsibilities of staff groups that are available and could be equally effective, sometimes even at lower cost. A potential example of this is nurse prescribing.[18] A step further is the creation of new workforce roles, compatible with the direction of reform. Community-based services may require a multi-disciplinary team involving a larger proportion of social workers and psychologists instead of only doctors and nurses. One could also consider new types of community workers focusing on prevention or early intervention, or support workers for patients with severe and enduring disorders. A primary care-based approach may suggest the development of mental health expertise for primary care workers. In practice, however, such models tend to create additional demands for staff and change roles of existing staff, rather than reducing the need for the specialists who remain in short supply.[19]

A short cut to solving staff shortages is international recruitment, setting up a three-way dilemma between ethical recruiting practice, the interest of both receiving and sending countries, and the benefit of individuals. In some trade zones the free migration of employees, including health care staff, is a basic right. It is also hard to argue that health care staff living in poverty and insecurity should be deprived of the opportunity to improve their lives by working in rich and secure countries. Developed countries argue that it is their responsibility to ensure access to health care for their population, although some consideration for the consequences of the country of origin should be shown. Ethical recruitment guidelines have been developed, distinguishing between active recruitment, that is deliberately setting out to attract staff, and passive recruitment, offering open opportunities, for example, by advertising in an international journal. The presence of a bilateral memorandum of agreement is also a indication of good practice.[20] It has to be recognized that international recruitment is not a phenomenon limited to western host countries. It is a global process, with a stream of staff following the trail of higher salaries.

Funding

If it is accepted that the public sector is responsible for the funding of mental health care, the total amount and by implication the proportion of the health budget to be allocated to mental health has to be determined, whether at a national, regional, or local level. In theory this could be calculated on the basis of a population needs assessment, matching the level of need with the cost of interventions. Many countries have performed epidemiological surveys with a sufficient degree of precision to allow an estimate of the prevalence of diagnostic groups.[21] It is more challenging to take account of the cost of treatment. Attempts to determine the costs of specific interventions are possible in research settings. However, in psychiatry, the cost variation even within homogeneous diagnostic groups attributable to factors such as severity, age, co-morbidity, ethnicity, geography, and social context, as well as supply factors such as the range of treatments and delivery options, make any prediction imprecise, as demonstrated by the challenge of developing Diagnosis Related Groups (DRGs) for mental health care.[22]

The variation of spending on mental health across countries is striking. About 20 per cent of countries in the world dedicate less than 1 per cent of the total health budget on mental health, and a

few more than 10 per cent as reported in the WHO Atlas project (2005),[23] with a strong positive correlation between the mental health budget and GDP. Thus, the poorest countries have a very low public spending on mental health, although this ignores the importance of out of pocket payments in many of these countries, sometimes with catastrophic impact on families.

Variations over time of the proportion of health budget allocated to mental health tend to be minor. This suggests that proportion of budget allocated to mental health can be attributed to static factors such as historical spending, often based on running costs of hospitals, adjusted by political priorities and influenced by factors such as advocacy and level of stigma. We do not know of any example of a national formula of rational health budget allocation at treasury/finance ministry level, although a number of countries use a psychiatric needs index to inform sub-national spending decisions.

Equitable resource allocation

A decision on basis of equity needs to balance the burden of disease for patients, families, and the community (see Chapter 7.5), intensity of human suffering, and the cost-effectiveness of treatment relative to other health conditions. In developed countries budgets can be expected to cover basic needs and simple evidence-based practices across all disease groups. It would be hard to accept in Western Europe or North America that patients have no access to treatments for conditions such as tuberculosis or schizophrenia. Some rationing in a variety of implicit or explicit forms is usually in place for expensive new technologies or costly treatments for common and mild conditions. Whilst mental health has few hi-tech interventions, interventions like CBT would be highly expensive if applied to more than defined subgroups of patients.

In poor countries with very low mental health budgets, many lives will depend on the decision how best to invest the scarce resources in order to gain the greatest cost-effectiveness as measured by DALYs averted. Cost-effectiveness studies have found that episodic treatment with older antidepressant drugs is the most cost-effective treatment in poor countries, whereas the treatment of schizophrenia is relatively more expensive.[24] Even more challenging is to decide on the merit of allocation across disease groups, for example, tuberculosis versus mental health.[25] However, one has to take into account that figures are based on averages, and ignore the very great personal and community benefits that some individual interventions can offer. In mental health, it also has to be considered that interventions are not only about treatment of diseases, but also involve in many instances action against major human rights abuses and the consequences of neglect, such as patients with schizophrenia locked up for years or children ignored in institutions. It is hard to put monetary values on such decisions, and this yet again places mental health activities beyond a narrow health treatment perspective, but shows it as one component of broad public policy making. The challenge is to keep mental health high on the agenda of all public policy makers, rather than it being allowed to slip down every government department's list of priorities.

Further information

Moniz, C. and Gorin, S.H. (2006). *Health and mental health care policy: a biopsychosocial perspective*. Allyn and Bacon, Boston.

Knapp, M., McDaid, D., Mossialos, E., *et al.* (2006). *Mental health policy and practice across Europe*. Open University Press, Maidenhead.

Brody, E. and Kemp, D.R. (eds.) (1993). *International handbook of mental health policy*. Greenwood Press, Westport.

Jenkins, R., McCulloch, A., Friedli, L., *et al.* (2002). *Developing a national mental health policy*. Maudsley Monograph. The Psychology Press, Hove.

References

1. Koontz, H. and Weihrich, H. (1988). *Management*. McGraw Hill, New York.
2. McCulloch, A. and Cohen, A. (2007). *Mental health policy and primary mental health care: present and future. Chapter in Mental health care policy and primary mental health care: present and future*. Royal College of General Practitioners, London.
3. Hughes, O.E. (1994). *Public management and administration*. St. Martin's Press, New York.
4. Peters, T. and Waterman, R. (1982). *In search of excellence: lessons from America's best Run Companies*. Harper and Row, New York.
5. Friedman, B. (2005). *The moral consequences of economic growth*. Oxford University Press, Oxford.
6. Bogdanor, V. (2005). *Joined up government*. Oxford University Press, Oxford.
7. Jenkins, R., McCulloch, A., Friedli, L., *et al.* (2002). *Developing a national mental health policy*. Maudsley Monograph 43. The Psychology Press, Hove.
8. Jablensky, A., Sartorius, N., Ensberg, G., *et al.* (1992). Schizophrenia: manifestations, incidence and cause in different cultures—a World Health Organization ten country study. *Psychological Medicine Monograph*, (Suppl. 20), 1–97.
9. WHO. (2005). *WHO resource book on mental health, human rights and legislation*. World Health Organization, Geneva.
10. WHO. (2005). *WHO resource book on mental health, human rights and legislation*. World Health Organization, Geneva.
11. WHO. (2001). *World Health Report 2001: mental health: new understanding, new hope*. World Health Organization, Geneva.
12. Beeharry, G., Whiteford, H., Chambers, D., *et al.* (2002). *Outlining the scope for public health sector involvement in mental health*. The World Bank, Washington, DC.
13. Offer, A. (2006). *The challenge of affluence*. Oxford University Press, Oxford.
14. Layard, R. (2006). *Happiness. Lessons from a new science*. Penguin Books, London.
15. Warr, P. (1987). *Work, unemployment and mental health*. Oxford University Press, Oxford.
16. WHO. (2003). *WHO mental health policy and service guidance package: organization of services for mental health*. World Health Organization, Geneva.
17. Muijen, M. and Ford, R. (1996). The market and mental health: intentional and unintentional incentives. *Journal of Interprofessional Care*, **10**, 13–22.
18. Department of Health. (2006). *From values to action: the chief officer's review of mental health nursing*. Department of Health, London.
19. Muijen, M. (2006). Challenges for psychiatry: delivering the mental health declaration for Europe. *World Psychiatry*, **5**, 113–17.
20. McIntosh, T., Torgerson, R., and Klassen, N. (2007). The ethical recruitment of internationally educated heath professionals: lessons from abroad and options for Canada. Canadian Policy Research Networks, Ottawa, Ontario.
21. WHO World Mental Health Survey Consortium. (2004). Prevalence, severity and unmet need for treatment of mental disorders in the

World Health Organization World Mental Health surveys. *The Journal of the American Medical Association*, **291**, 2581–9.
22. Burgess, P., Pirkis, J., Buckingham, W., *et al.* (1999). Developing a casemix classification for specialist mental health services. *Casemix*, **4**.
23. WHO. (2005). *Atlas: mental health resources in the world 2005*. World Health Organization, Geneva.
24. WHO. (2006). *Dollars, DALYs and decisions: economic aspects of the mental health system*. World Health Organization, Geneva.
25. Jamison, D.T., Breman, J.G., Measham, A.R., *et al.* (2006). *Disease control priorities in developing countries* (2nd edn). World Bank/Oxford University Press, New York.

7.2

Service needs of individuals and populations

Mike Slade, Michele Tansella, and Graham Thornicroft

Introduction

The importance of needs assessment has been one of the most consistent themes to emerge from the evolution of community mental health services. However, the concept of 'need' is used in different, and sometimes contradictory, ways. The aim of this chapter is to

- define needs assessment
- consider different approaches to assessing needs, both at the individual and at the population levels
- discuss how needs assessments can be applied in real-world settings in planning and delivering clinical care.

Defining needs

At its simplest, a need involves a lack of something. But of what? In operational terms the concept of need is usually applied to a difficulty (in this case in relation to a person with mental illness) for which a possibly effective intervention exists. By implication an experienced difficulty for which there is no known effective intervention is therefore not defined as a need.[1] The clearest categorization of such needs was identified by Brewin, who grouped definitions of need within mental health care into three categories: lack of health, lack of access to services or institutions, and lack of action by mental health workers.[2] Approaches to need within each of these three categories will be reviewed.

(a) Needs for improved health

The psychologist Maslow established probably the best-known hierarchy of need, when he formulated a theory of human motivation.[3] In his model, fundamental physiological needs (such as the need for food) underpin the higher needs of safety, love, self-esteem, and self-actualization. He proposed that people are motivated by the requirement to meet these needs, and that higher level needs could only be met once the lower and more fundamental needs were met. The clinical relevance of this theory is that it implies a hierarchy of clinical priorities—interventions to meet basic physiological need (e.g. to ensure adequate food supply) should take priority over interventions to foster, for example, self-esteem.

In practice health-related needs are often considered in a widely defined way. In England, for example, the requirement to base the provision of services on level of need was first made explicit in the National Health Service and Community Care Act,[4] which defined need as *the requirements of individuals to enable them to achieve, maintain, or restore an acceptable level of social independence or quality of life*. This requirement was retained when national standards for mental health services were set.[5] This involves **needs-led** care planning—basing care for an individual patient on an assessment of their health and social needs. The needs-led approach offers many benefits:

1 The overall level of need gives guidance about which part of the mental health system should treat the patient, for example that people with less disabling mental disorders should be seen in primary care settings.[6]

2 Needs assessment can improve the comprehensiveness of case formulations and care plans by incorporating a broad range of health determinants, such as poor housing or lack of social support.

3 Explicit identification of need can support clinician–patient discussions about care priorities, which is associated with improved treatment satisfaction[7,8] and compliance.[9,10]

4 Identification of needs helps to identify the contribution of services outside the psychiatric sector.

5 Needs-led care can facilitate more individualized treatment planning than diagnosis-driven approaches, by more closely matching the help offered to patient's needs and by explicitly identifying problems which require the involvement of both health and other agencies.

Needs-led care planning focussing on health can be differentiated from the assessment of care needs. Assessing care needs involves identifying whether the patient will benefit from a predefined menu of interventions, and by definition will not identify all unmet needs for individual patients. Assessment of need at the patient level should therefore be a separate process from decisions about what care or treatment to provide. There are, however, other reasons to assess needs for services, which we now review.

(b) Needs for services

The second category of need is a requirement for a particular type of service. At the population level, it is possible to use epidemiological methods to develop prevalence for different disorders, which can be translated into estimates of the need for services. A recent epidemiological survey in the United States, for example, found very considerable unmet need of the population level nationwide.[11] This study identified that between 1990–92 and 2001–03 the overall annual period prevalence of mental illnesses remained constant at between 29.4 and 30.5 per cent. Among these cases, however, there was an increase in the proportion who received any treatment at all, rising from 20.3 to 32.9 per cent between the two time periods. The inverse is however very revealing, namely that the most recent data show that 67 per cent of people with mental disorders in the United States receive no treatment. The situation is worse in other countries. A recent comparative international study of depression found that 0 per cent of patients in St Petersburg received evidence-based treatment in primary care, and only 3 per cent were referred on to specialist mental health care.[12] The inability of patients to afford out-of-pocket costs was the primary barrier to care for 75 per cent of the depressed Russian patients studied.

International comparisons of population-level needs have been conducted in recent years. The ESEMed Study, for example, carried out cross-sectional surveys in Belgium, France, Germany, Italy, the Netherlands, and Spain among 8796 representative members of the general population. Individuals with a 12-month mental disorder that was disabling or that had led to use of services in the previous 12 months were considered in need of care. The study found that about 6 per cent of the sample was defined as being in need of mental health care. Nearly half (48 per cent) of these people reported no formal health care use, so that 3.1 per cent of the adult population had an unmet need for mental health care. In contrast, only 8 per cent of the people with diabetes had reported no use of services for their physical condition.[13]

(c) Needs for action

In health care, the concept of need has been taken to mean the ability to benefit in some way from health care, and thus distinguished from demand (what the person asks for) and supply (services given).[14] For example, the MRC Needs for Care Assessment Schedule is premised on the assumption that need is 'a normative concept which is to be defined by experts'.[15]

Using this approach, an Australian study compared current and optimal treatment for 10 high-burden mental disorders in Australia.[16] This found that current levels of treatment at current coverage avert 13 per cent of the overall burden attributable to these disorders. Providing optimal treatment at current coverage would avert 20 per cent of the burden, and optimal treatment at optimal coverage would avert 28 per cent. The development of a more robust treatment evidence base makes this innovative approach to informing public policy more possible, and the approach can be recommended for evidence-based policy initiatives.

Patient and staff perceptions of need

There has been a long-standing recognition that differences in perceptions of need can exist, in particular between staff and patient. In the 1990s the emphasis was put on acknowledging these differences, but then prioritizing the staff perspective. For example, UK policy stated that *all users … should be encouraged to participate to the limit of their capacity. …Where it is impossible to reconcile different perceptions, these differences should be acknowledged and recorded*.[17] Several societal and scientific developments challenge this prioritization of staff over patient perspectives.

First, general societal changes towards consumerism and an emphasis on rights have produced more assertive mental health service users. Easier access by patients to internet-based information reduces the knowledge disparity. Reduced societal trust in the authoritative expert has eroded the position power of mental health staff. The emphasis put on choice and empowerment raise patient expectations of being more than passive recipients of care.[18,19]

Second, the prioritization of staff perspectives has been actively challenged by an increasingly vociferous and organized user movement. This opposition has found its voice in the 'recovery' movement, which emphasizes the meaning and values of the patient, and the need for services to foster self-management rather than dependency. There has been widespread international policy support for recovery-focussed services[20] although there can be tensions between what professionals construe as their duty of care and being led by the patient perspective on need, which can create ethical dilemmas. Care planning which emphasizes agreement between staff and patients may have additional advantages. A recent study in Verona showed staff–patient agreement on needs was significantly associated with better treatment outcomes both rated by the patient and by staff (psychopathology, social disability, global functioning, subjective quality of life, and satisfaction with care).[21] Similarly, there is emerging evidence that crisis plans (advanced statements) which are jointly agreed between staff and patient can be cost-effective in reducing compulsory admission to hospital.[22,23] Such emerging findings indicate that needs assessment and care planning, which are based on negotiation and jointly agreed analyses of problems and interventions, are likely to become increasingly important in future.

Finally, emerging empirical evidence strongly supports the positioning of the patient perspective at the heart of needs assessment and care planning. Evidence from several studies consistently shows differences between staff and patient perspectives on need,[24,25] so the two perspectives are not interchangeable. Empirical research suggests two reasons for basing care on the patient rather than staff assessment of need. First the patient rating is more stable than the staff rating.[26] Second, longitudinal studies indicate a causal relationship between patient-rated (but not staff-rated) unmet need and quality of life.[27–29] If the goal of mental health services is to improve quality of life, then best available evidence indicates that the patient's perspective on their unmet needs should drive care planning.

Assessing needs

In this section we identify specific approaches to assessing needs.

(a) Individual-level needs assessment measures

Several standardized approaches to the assessment of patient-level need have been developed, primarily in the United Kingdom. These have shown a transition along a continuum, from an initial focus on assessment of need as an objective state to be defined by experts following careful assessment, towards those which emphasize the subjective nature of needs assessment.

The earliest standardized needs assessment measure was the **Medical Research Council Needs for Care Assessment (NFCAS)**.[30]

The NFCAS assesses the need for further action by health care professionals, and links identification of a need with a predefined list of actions. This raises two problems. First, the emphasis on identifying available interventions which would be at least partly effective is problematic, given the complexities of deciding that a treatment has not worked. Second, updating the list of actions has proved problematic. However, as Bebbington notes, 'the inevitable value judgements inherent in the procedure have the virtue of being public and consequently accessible to argument'.[31] An important variation of the NFCAS is the **Cardinal Needs Schedule (CNS)**,[32] which also considers patient willingness to accept help and level of carer concern. Training is needed for using both the NFCAS and the CNS, and they are primarily used for research purposes.

At the other end of this continuum are needs assessment measures which emphasize individual difference and the subjective nature of need. The **AVON Mental Health Measure** was developed by service users, and assesses physical, social, behaviour, access, and mental health domains.[33] It can take up to 20 min for completion by the patient and 5 min by the staff, and its development has emphasized external validity over other psychometric properties. The **Carers and Users Experience of Services (CUES)** was developed by service users and staff, and assesses 16 domains: the place you live, money situation, the help you get, the way you spend time, your relationships, social life, information/advice, access to services, choice of mental health services, relationship with mental health workers, consultation and contact, advocacy, stigma, any treatment, access to physical health services, and relationship with physical health workers.[34] Completion can take up to 30 min. Neither AVON nor CUES have become widely used in mental health services.

The **Camberwell Assessment of Need (CAN)**[35] spans both ends of the continuum. It assesses 22 domains of health and social need, and a key development is that it records staff and patient views separately, without giving primacy to either perspective. Research (CAN-R), clinical (CAN-C), and brief versions (CANSAS) of the CAN have been developed for adults of working age with severe mental health problems,[36] and it has been translated in 22 languages. Variants have been developed for people with learning disabilities and mental health problems (CANDID),[37] mentally disordered offenders (CANFOR),[38] older adults (CANE),[39] and mothers with mental health problems (CAN-M).[40] An updated web resource for the CAN is available at www.iop.kcl.ac.uk/prism/can.

The CAN has become the most widely used needs assessment measure internationally,[41] and is the standardized needs assessment measure which is most relevant to routine clinical practice. The short version, CANSAS, can be recommended for routine use in community services. Two specific approaches have been empirically shown to produce patient-level benefit. First, the patient-rated two-way communication (2-COM) measure is an amended version of the CAN which gives the patient the opportunity to identify unmet needs and also prioritize those which they wish to discuss with their clinician.[42] Asking patients to complete 2-COM before an outpatient appointment and then using that information in the appointment was associated with greater patient satisfaction and more likelihood of treatment change.[7] Second, a structured approach to collating and feeding back staff and patient ratings for CANSAS and other assessments led to a reduction in psychiatric admissions, probably because of earlier intervention during relapse,[43] and improvements in patient-rated unmet need and quality of life for higher premorbid IQ patients.[44] Routine use of CANSAS brings patient-level benefits, and empirical evidence indicates the clinical focus should be on assessment and interventions for patient-rated, rather than staff-rated, unmet needs.

(b) Population-level needs assessment measure

Measures to assess population-based needs can be classified by the data and by the analytic approaches they use. Three types of data are commonly used. The most readily accessible documents the use of current mental health services. While this can be criticized as reflecting only current service provision, its ready availability and nationwide coverage means that is extensively used.

Simple population-based samples are very inefficient in estimating the prevalence of relatively rare conditions such as schizophrenia. Studies thus tend to use two-stage procedures, with a relatively brief initial screening process applied to a large number of people followed in-depth interviews for a selected few. 'Booster' samples, perhaps including all the known psychiatric patients for the areas surveyed, may be sought through mental health services.[45] Population surveys depend on the identification of randomly sampled individuals. Some types of mental health problems, notably substance misuse, are commonly associated with socially marginal lifestyles, making it likely that sufferers will be systematically underrepresented by traditional population sampling approaches. More sophisticated sampling approaches, such as capture–recapture methods, have been used for these situations.[46]

The third type of data relates to the views of local people. Local needs assessment studies entail a structured approach to eliciting the views of service users, their carers, interested voluntary sector organizations, and all statutory agencies with responsibilities in the area. Smith[47] has described how this type of study can be integrated into the overall planning process.

Government initiatives in England have tried to base the allocation of money between areas on the morbidity as well as the size of their populations. This has led to studies modelling this variation. The first widely used index[48] was developed on the basis of consensus between GPs about patient characteristics associated with high use of primary care services. While developed for wider purposes, this was shown to relate reasonably closely to variations in psychiatric admission. Later indices have been established by statistical modelling exercises seeking to quantify the relationship between social variables measured in censuses and either service use,[49] or population-based epidemiological findings. The variation between places in the prevalence of the less severe types of mental illness commonly dealt with in primary care is less than that for problems usually managed by specialist mental health services, which again is much less than that observed for forensic services. Thus models developed for one level of care should not be used to estimate patterns of need for other levels.

In practice, no single approach to assessing the needs of a population will suffice. Needs assessment at this level requires the integration of many perspectives. The Kings Fund review of London's mental health Services[50] illustrates how a detailed perspective can be assembled from many fragments of evidence, each of which would be inadequate in isolation. Recent examples of population-level needs assessment from the United States, Canada, and

New Zealand also reveal that epidemiological studies may not produce data that corresponds directly to needs, and that some sub-populations, for example particular ethnic groups, may be less well represented in such approaches, unless considerable methodological care is taken.[51–53]

If a mental health practitioner is asked to join a committee to plan services, for example, for a local catchment area, what approach is helpful to identify and use population-level needs? Table 7.2.1 indicates a series of steps to find the best available information on population prevalence rates.[54]

As Table 7.2.1 shows, we consider that the best possible information would be local epidemiological data on the occurrence of mental disorders, using a standard system of classification, alongside a measure of the needs for treatment among the prevalent cases identified.[55] Since these assessments are expensive and time-consuming, most sites will not have access to such recent local data. If the data in step (1) are not available then we suggest that country/regional epidemiological data (2) are used instead, and are then weighed for local socio-demographic characteristics. But if such larger scale prevalence data are not available, then a third option is to use international rates from 'comparison' countries or regions, again weighted for local socio-demographic characteristics (3). The results in this case will be less accurate because they are based on the additional assumption that the data can be transferred between countries. A newer set of techniques that offer considerable promise are rapid appraisal/rapid assessment techniques. These are methods to undertake brief assessments of population needs which are focussed upon key focussed questions, for example on how primary care services should be augmented to treat people with depression, and example of these approaches have been used to positive effect in South Africa.[56–58]

In some cases, none of the data described in steps 1–3 will be available, and then the next option (4) is to use a number of experts, some of whom may be from the local area, to produce a consensus statement on the local rates and characteristics of people with mental illness. Such a data synthesis can be based on the best available views, taking into account local factors (e.g. levels of non-health service provision, family support, traditions, degree of affluence, or migration). This pragmatic approach will yield data which are accurate enough to use for local service planning purposes.

Table 7.2.1 Ways to measure or estimate local population mental health prevalence

(1) Actual local epidemiological data on psychiatric morbidity and disability for the particular area by age, sex, ethnicity, social status, and degree of urbanicity
(if not available) ↓
(2) Country/regional epidemiological data weighed for local socio-demographic characteristics
(if not available) ↓
(3) International data from 'comparable' countries or regions, adjusted for local socio-demographic characteristics
(If 1, 2, 3 not sufficient) ↓
(4) Best estimates and expert synthesis and interpretation based on other sources of local information and opinions (e.g. extent of non-health service provision, family support, local traditions, or migration)

This raises the important topics of coverage and focussing. *Coverage* means the proportion of people who receive treatment who could benefit from it.[59] *Focussing* refers to how far those people who actually receive treatment in fact need it: do they have any form of mental illness?[60] Even in the most well-resourced countries one can find both low coverage and poor focussing.[61,62] From the public health perspective, therefore, the key issue is the appropriate use of resources, whatever the level of resources actually available, namely to increase both coverage and focus.

(c) The relationship between individual and population-level needs

We have argued in this chapter that the provision of care for individual patients should be based on assessment of their health and social needs. Can these individual assessments be aggregated to inform service? For population-level service planning, the key question is what types of interventions to provide, and with what capacity. Therefore data from individual needs assessments cannot simply be aggregated to inform service development decisions. While it is theoretically feasible to undertake routine standardized needs assessments on all patients within a service, this approach alone has three drawbacks. Firstly, despite the developing evidence reviewed earlier about the benefits of routine use of standardized outcome assessments, this remains the exception rather than the norm.[63,64] Secondly, there is not yet a sufficiently developed information infrastructure to support the national collection, management, and analysis of such data, despite this being a priority identified two decades ago.[65] Thirdly, even if individual needs assessments were nationally aggregated, the diversity of views (patient, staff, carer, taxpayer) would make a shared interpretation of the data problematic.

Conclusion

In this chapter we have emphasized that it is of central importance when planning mental health service for populations, to do so on the basis of (i) the occurrence of mental disorders in that particular population, (ii) the impairments caused by these disorders that require interventions, (iii) the nature and level of needs among these people, (iv) identifying from among these needs those which are unmet, and then (v) prioritizing new service development on the basis of these unmet needs, including a range of social supports and services (such as housing or employment opportunities, outside the mental health system), the requirements for enhanced physical/general health care, as well as improvements in the provision of specific mental health services. For all of these sectors there is an increasingly clear call from service user/consumer groups for involvement in these priority-setting planning exercises.[18,19]

At the level of individuals with mental illness, there is a similar trend to increasingly involve service users/consumers in assessing needs, with emerging evidence that this produces a more comprehensive basis for care planning. Indeed in the last decade there has been an important conceptual shift away from the view that professionals defined 'needs' while consumers stated 'demands', to a better appreciation of the many advantages to be gained from identifying, as far as possible, unmet needs in a joint and consensual way as a basis for action.

Further information

Andrews, G. and Henderson, S. (eds.) (2000). *Unmet need in psychiatry*. Cambridge University Press, Cambridge.

Lasalvia, A. and Ruggeri, M. (2007). Multidimensional outcomes in 'real world' mental health services: follow-up findings from the South Verona Project. *Acta Psychiatrica Scandinavica Supplementum*, **437**(S116), 3–77.

Thornicroft, G. (2001). *Measuring mental health needs* (2nd edn). Gaskell, Royal College of Psychiatrist, London.

Thornicroft, G., Becker, T., Knapp, M., *et al.* (2006). *International outcome measures in mental health. Quality of life, needs, service satisfaction, costs and impact on carers*. Gaskell, Royal College of Psychiatrists, London.

Thornicroft, G. and Tansella, M. (2008). *Better mental health care*. Cambridge University Press, Cambridge.

Regularly updated web site for the Camberwell Assessment of Need is available at: http://www.iop.kcl.ac.uk/prism/can

References

1. Brewin, C.R. (2001). Measuring individual needs for care and services. In *Measuring mental health needs* (2nd edn) (ed. G. Thornicroft), pp. 273–90. Gaskell, London.
2. Brewin, C. (1992). Measuring individual needs for care and services. In *Measuring mental health needs* (eds. G. Thornicroft, C. Brewin, and J. Wing). Gaskell, Royal College of Psychiatrists, London.
3. Maslow, A. (1954). *Motivation and personality*. Harper & Row, New York.
4. House of Commons (1990). *National health service and community care act*. HMSO, London.
5. Department of Health. (1999). *Mental health national service framework*. HMSO, London.
6. Tansella, M. and Thornicroft, G. (1999). *Common mental disorders in primary care*. Routledge, London.
7. van Os, J., Altamura, A.C., Bobes, J., *et al.* (2004). Evaluation of the two-way communication checklist as a clinical intervention. *The British Journal of Psychiatry*, **184**, 79–83.
8. Lasalvia, A., Bonetto, C., Malchiodi, F., *et al.* (2005). Listening to patients' needs to improve their subjective quality of life. *Psychological Medicine*, **35**, 1655–65.
9. Haynes, R.B., McDonald, H.P., and Garg, A.X. (2002). *Interventions for helping patients to follow prescriptions for medications*. Cochrane Library Update Software, Oxford.
10. Gray, R., Leese, M., Bindman, J., *et al.* (2006). Adherence therapy for people with schizophrenia. European multicentre randomised controlled trial. *The British Journal of Psychiatry*, **189**, 508–14.
11. Kessler, R.C., Demler, O., Frank, R.G., *et al.* (2005). Prevalence and treatment of mental disorders, 1990 to 2003. *The New England Journal of Medicine*, **352**, 2515–23.
12. Simon, G.E., Fleck, M., Lucas, R., *et al.* (2004). Prevalence and predictors of depression treatment in an international primary care study. *The American Journal of Psychiatry*, **161**, 1626–34.
13. Alonso, J., Codony, M., Kovess, V., *et al.* (2007). Population level of unmet need for mental healthcare in Europe. *The British Journal of Psychiatry*, **190**, 299–306.
14. Stevens, A. and Gabbay, J. (1991). Needs assessment needs assessment. *Health Trends*, **23**, 20–3.
15. Bebbington, P. and Rees, S. (2001). Assessing the need for psychiatric services at the district level: using the results of community surveys. In *Measuring mental health needs* (2nd edn) (ed. G. Thornicroft). Gaskell, Royal College of Psychiatrist, London.
16. Andrews, G., Issakidis, C., Sanderson, *et al.* (2004). Utilising survey data to inform public policy: comparison of the cost-effectiveness of treatment of ten mental disorders. *The British Journal of Psychiatry*, **184**, 526–33.
17. Department of Health Social Services Inspectorate. (1991). *Care management and assessment: practitioners' guide*. HMSO, London.
18. Chamberlin, J. (2005). User/consumer involvement in mental health service delivery. *Epidemiologia e Psichiatria Sociale*, **14**, 10–4.
19. Rose, D., Thornicroft, G., and Slade, M. (2006). Who decides what evidence is? Developing a multiple perspectives paradigm in mental health. *Acta Psychiatrica Scandinavica Supplementum*, **429**(S113) 109–14.
20. New Freedom Commission on Mental Health. (2005). *Achieving the promise: transforming mental health care in America*. Department of Health and Human Services, Rockville, MD.
21. Lasalvia, A., Bonetto, C., Tansella, M., *et al.* (2007). Does staff-patient agreement on needs for care predict a better mental health outcome? A 4 year follow-up in a community service. *Psychological Medicine*, doi 10.1017/S0033291707000785.
22. Henderson, C., Flood, C., Leese, M., *et al.* (2004). Effect of joint crisis plans on use of compulsory treatment in psychiatry: single blind randomised controlled trial. *British Medical Journal*, **329**, 136.
23. Flood, C., Byford, S., Henderson, C., *et al.* (2006). Joint crisis plans for people with psychosis: economic evaluation of a randomised controlled trial. *British Medical Journal*, **333**, 729.
24. Lasalvia, A., Ruggeri, M., Mazzi, M.A., *et al.* (2000). The perception of needs for care in staff and patients in community-based mental health services. The South-Verona Outcome Project 3. *Acta Psychiatrica Scandinavica*, **102**, 366–75.
25. Hansson, L., Vinding, H.R., Mackeprang, T., *et al.* (2001). Comparison of key worker and patient assessment of needs in schizophrenic patients living in the community: a Nordic multicentre study. *Acta Psychiatrica Scandinavica*, **103**, 45–51.
26. Slade, M., Leese, M., Taylor, R., *et al.* (1999). The association between needs and quality of life in an epidemiologically representative sample of people with psychosis. *Acta Psychiatrica Scandinavica*, **100**, 149–57.
27. Slade, M., Leese, M., Ruggeri, M., *et al.* (2004). Does meeting needs improve quality of life? *Psychotherapy and Psychosomatics*, **73**, 183–9.
28. Lasalvia, A., Bonetto, C., Malchiodi, F., *et al.* (2005). Listening to patients' needs to improve their subjective quality of life. *Psychological Medicine*, **35**, 1655–65.
29. Slade, M., Leese, M., Cahill, S., *et al.* (2005). Patient-rated mental health needs and quality of life improvement, *The British Journal of Psychiatry*, **187**, 256–61.
30. Brewin, C., Wing, J., Mangen, S., *et al.* (1987). Principles and practice of measuring needs in the long-term mentally ill: the MRC needs for care assessment. *Psychological Medicine*, **17**, 971–81.
31. Bebbington, P. (1992). Assessing the need for psychiatric treatment at the district level: the role of surveys. In *Measuring mental health needs* (eds. G. Thornicroft, C. Brewin, and J. Wing). Gaskell, Royal College of Psychiatrists, London.
32. Marshall, M. (1994). How should we measure need? *Philosophy, Psychiatry and Psychology*, **1**, 27–36.
33. Lelliott, P. (2000). What do people want from specialist mental health services and can this be routinely measured in routine service settings? *Behavioural and Cognitive Psychotherapy*, **28**, 361–8.
34. Lelliott, P., Beevor, A., Hogman, J., *et al.* (2001). Carers' and users' expectations of services—user version (CUES-U): a new instrument to measure the experience of users of mental health services. *The British Journal of Psychiatry*, **179**, 67–72.
35. Phelan, M., Slade, M., Thornicroft, G., *et al.* (1995). The Camberwell assessment of need: the validity and reliability of an instrument to assess the needs of people with severe mental illness. *The British Journal of Psychiatry*, **167**, 589–95.
36. Slade, M., Loftus, L., Phelan, M., *et al.* (1999). *The Camberwell assessment of need*. Gaskell, London.

37. Xenitidis, K., Slade, M., Bouras, N., *et al.* (2003). *CANDID: Camberwell assessment of need for adults with developmental and intellectual disabilities*. Gaskell, London.
38. Thomas, S., Harty, M., Parott, J., *et al.* (2003). *The forensic CAN: Camberwell assessment of need forensic version (CANFOR)*. Gaskell, London.
39. Reynolds, T., Thornicroft, G., Abas, M., *et al.* (2000). Camberwell assessment of need for the elderly (CANE). Development, validity and reliability. *The British Journal of Psychiatry*, **176**, 444–52.
40. Howard, L., Slade, M., O'Keane, V., *et al.* (eds.) (2008). *The Camberwell assessment of need for pregnant women and mothers with severe mental illness*. Gaskell, London.
41. Evans, S., Greenhalgh, J., and Connelly, J. (2000). Selecting a mental health needs assessment scale: guidance on the critical appraisal of standardized measures. *Journal of Evaluation in Clinical Practice*, **6**, 379–93.
42. van Os, J., Altamura, A.C., Bobes, J., *et al.* (2002). 2-COM: an instrument to facilitate patient-professional communication in routine clinical practice, *Acta Psychiatrica Scandinavica*, **106**, 446–52.
43. Slade, M., McCrone, P., Kuipers, E., *et al.* (2006). Use of standardised outcome measures in adult mental health services: randomised controlled trial. *The British Journal of Psychiatry*, **189**, 330–6.
44. Slade, M., Leese, M., Gillard, M., *et al.* (2006). Premorbid IQ and response to routine outcome assessment. *Psychological Medicine*, **36**, 1183–92.
45. Jenkins, R., Bebbington, P., Brugha, T., *et al.* (1997). The National Psychiatric Morbidity surveys of Great Britain-strategy and methods. *Psychological Medicine*, **27**, 765–74.
46. Hay, G. and McKeganey, N. (1996). Estimating the prevalence of drug misuse in Dundee, Scotland: an application of capture-recapture methods. *Journal of Epidemiology and Community Health*, **50**, 469–72.
47. Smith, H. (1998). Needs assessment in mental health services: the DISC framework. *Journal of Public Health Medicine*, **20**, 154–60.
48. Jarman, B. (1983). Identification of underprivileged areas. *British Medical Journal (Clinical research ed.)*, **286**, 1705–9.
49. McCrone, P., Thornicroft, G., Boyle, S., *et al.* (2006). The development of a local index of need (LIN) and its use to explain variations in social services expenditure on mental health care in England. *Health & Social Care in the Community*, **14**, 254–63.
50. Johnson, S., Ramsay, R., Thornicroft, G., *et al.* (1997). *London's mental health. The report of the Kings Fund Commission*. Kings Fund Publishing, London.
51. Messias, E., Eaton, W., Nestadt, G., *et al.* (2007). Psychiatrists' ascertained treatment needs for mental disorders in a population-based sample. *Psychiatric Services*, **58**, 373–7.
52. Kumar, S., Tse, S., Fernando, A., *et al.* (2006). Epidemiological studies on mental health needs of Asian population in New Zealand. *The International Journal of Social Psychiatry*, **52**, 408–12.
53. Hanson, L., Houde, D., McDowell, M., *et al.* (2006). A population-based needs assessment for mental health services. *Administration and Policy in Mental Health*, **34,** 233–42.
54. Thornicroft, G. and Tansella, M. (1999). *The mental health matrix: a manual to improve services*. Cambridge University Press, Cambridge.
55. Thornicroft, G. (2001). *Measuring mental health needs* (2nd edn). Gaskell, Royal College of Psychiatrists, London.
56. Flisher, A.J., Lund, C., Funk, M., *et al.* (2007). Mental health policy development and implementation in four African countries. *Journal of Health Psychology*, **12**, 505–16.
57. Lund, C. and Flisher, A.J. (2006). Norms for mental health services in South Africa. *Social Psychiatry and Psychiatric Epidemiology*, **41**, 587–94.
58. Lund, C. (2002). *Mental health service norms in South Africa*. PhD thesis. University of Cape Town, Cape Town.
59. Habicht, J.P., Mason, J.P., and Tabatabai, H. (1984). Basic concepts for the design of evaluations during programme implementation. In *Methods for the evaluation of the impact of food and nutrition programmes. Food and nutrition bulletin* (eds. D.R. Sahn, R. Lockwood, and N.S. Scrimshaw), pp. 1–25. The United Nations University, New York.
60. Tansella, M. (2006). Recent advances in depression. Where are we going? *Epidemiologia e Psichiatria Sociale*, **15**, 1–3.
61. World Health Organisation (2005). *Mental health action plan for Europe.* World Health Organisation, Copenhagen.
62. World Health Organisation (2005). *Mental health declaration for Europe.* World Health Organisation, Copenhagen.
63. Gilbody, S.M., House, A.O., and Sheldon, T.A. (2002). Psychiatrists in the UK do not use outcome measures. *The British Journal of Psychiatry*, **180**, 101–3.
64. Valenstein, M., Mitchinson, A., Ronis, D.L., *et al.* (2004). Quality indicators and monitoring of mental health services: what do frontline providers think? *The American Journal of Psychiatry*, **161**, 146–53.
65. Ellwood, P. (1988). Outcomes management–a technology of patient experience. *The New England Journal of Medicine*, **318**, 1549–56.

7.3

Cultural differences care pathways, service use, and outcome

Jim van Os and Kwame McKenzie

This chapter discusses the influence of culture on the route an individual takes to access treatment for psychological distress and the treatment received. Culture is difficult to measure. All categories of cultural variables have different meanings and measure different things. Research into them leads to different hypotheses. Given this reality, there is no need to join the cul-de-sac argument of whether one or the other is the most important. In the discussion below, the roles of ethnicity and of socio-economic, political, community, national, and other factors that help to define the culture of an individual or a group are acknowledged. Associations with pathways to care, service use, and outcome will be presented.

Care pathways

Cultural variation in pathways through care

At the beginning of the pathway to care, the individual displays cognitive, physical, or behavioural changes. They or their family, friends, or wider community interpret these as in need of some remedy. The individual's personal resources and then informal resources of family and friends are often triggered to help deal with the problem. These may lead to resolution but if they do not they may lead to presentation through an ever more distant and 'professional' array of caregivers, help agencies, and formal medical services. Most help for psychological problems is not given by mental health services. Interventions and their perceived success or failure move an individual along a pathway.

Pathways through care have differing directions and durations. These depend on where the pathway starts, the presenting symptoms, and psychosocial and cultural factors in the individual, their community, and the services used. Pathways are not random, they are structured and set by a dialogue between the individual, the community, and the code of the statutory services and the law set within that country.[1]

At each level the aim of care is to help the individual move down to less professional interventions until they are either back in the community or at the lowest intensity of care that meets their needs. Traditionally, care pathways have considered routes to getting treatment but with de-institutionalization, social inclusion and the recovery model, pathways out of care need to be considered more widely.

International comparisons

Cultural variation in pathways to care for a mental health problem is readily, though crudely, demonstrated through international comparisons. For example, in an international study of the pathways to care of 1554 patients newly referred to the mental health services in 11 centres in different countries, the majority of patients (63–80 per cent) were referred by their general practitioner in United Kingdom, Spain, Portugal, Czechoslovakia, Cuba, Mexico, and Aden Democratic Republic of Yemen. Only between 0–15 per cent of patients in these centres had referred themselves, with the exception of Mexico (24 per cent). In Kenya, only 7 per cent were referred by general practitioners but 72 per cent were referred by a hospital doctor. In Pakistan and India, a quarter of patients were referred by general practitioners, a quarter by hospital doctors, a third were self-referred, and 11–17 per cent were referred by religious healers. In Indonesia, both primary care and native healer referrals constituted each around a third of all referrals. The differences between these 11 centres largely reflected the people's choice of first port of call for a psychiatric problem.[2]

International differences in pathways are also reflected in primary health care studies. In a 14 country World Health Organization (WHO) investigation, the proportion of attendees with anxiety and depression as defined by the Composite International Diagnostic Interview (CIDI) varied five-fold across centers. Asian sites reported the lowest rates and European and South American sites the highest. The differences in prevalence may reflect a combination of demographic differences between attendees, true differences in population prevalence, the differential availability of other culture-specific pathways to care for the psychologically distressed, and differential sensitivity of the CIDI in picking up psychiatric disorder in different cultures.[3]

Geographically less dispersed countries also demonstrate significant differences in care pathways. In a recent study of 6 Eastern European countries, the percentage of new patients with schizophrenia who first sought care from psychiatric services ranged

from 69 per cent in Bucharest (Romania) to 47 per cent in Zagreb (Croatia). Thirteen percent of patients first sought care from general practitioners in Strumica, Macedonia but 47 per cent in Zagreb. The police were the first port of call in 8 per cent of cases in Bucharest but for none in Zagreb.[4]

This study used the 'encounter form' developed for the WHO which can be used to map and quantify pathways to care (Fig. 7.3.1).

Marked differences in pathways to care also exist within countries and groups. The increase in the use of involuntary admission of African-Caribbeans in the UK and African-Americans in the US is well documented. The reasons for this are unclear but some argue that it reflects service configuration.[5] More recently, increased involuntary admission rates have been reported in the Maori population of Auckland, New Zealand.[6]

Geographic and ecological factors may also contribute to variation in pathways to care within a country. In a Canadian province, involuntary admissions were shown to be related to the size of a community and its proximity to the hospital. Thus, involuntary admission rates were increased if a community was close to the hospital. Rates were also higher in both densely populated inner city areas as well as in cities of less than 500 people. The higher rates of involuntary admission in small towns may be related to a decreased likelihood of being tolerated or remaining anonymous.[7] Despite a similar mental health act, involuntary admissions in Greenland were found to be twice as high as they were in Denmark or the Faroe Islands. The excess risk was associated with the higher homicide rate, lower psychiatric bed availability, lower access to psychiatric care, small settlements, and increased alcohol consumption and violence in Greenland.[8]

Factors associated with cultural variation in pathways

How interpersonal or cultural factors are translated into differences in pathways has rarely been assessed. However, a number of factors have been identified that contribute to cross-national and cross-ethnic differences.

The family may play an important role. An American study reported that Chinese patients were kept for extended periods of time within their families at the beginning of pathways, while Anglo-Saxons and Central Europeans were referred by their relatives or themselves to a range of mental health and social agencies. Native Americans tended to be referred by people other than relatives or themselves.[9] In another study, both Asians and African-Americans showed more extended family involvement, and the involvement of key family members tended to be persistent and intensive in Asians. Ethnicity was also associated with the length of delay, Asians showing the longest delay and white people, the shortest.[10]

The history of and the way that institutions promote themselves can affect the attitude minority patients have to them and so their likelihood of using them. For example, the view of American hospital services by ethnic minority patients has been tarnished by the American Medical Association's support for segregated wards until the 1960s.[11]

The experience of illness is a culturally shaped phenomenon. The monitoring of change, the understanding of symptoms, the language used to present symptoms, and the fears that accompany symptoms are all suffused with cultural interpretations.[12] Cultural differences in displaying distress are most obviously seen in culture-bound syndromes but are present in the content of delusions and somatic presentation of distress. Somatic symptoms are located in multiple systems of meaning that serve diverse psychological and social functions. In one study, the experience of neurotic patients in India was labeled depressive by clinicians using DSM-IIIR, whereas the patients emphasized their somatic experience.[13] Such discrepancies between professional theory and patients' experience may have an impact on the recognition and treatment of psychiatric disorder. For example, ethnic difference between the doctor and the patient or linguistic/communication problems had previously been offered as reasons for British general practitioners missing depression in South Asian women. However, South Asian doctors are also more likely to miss

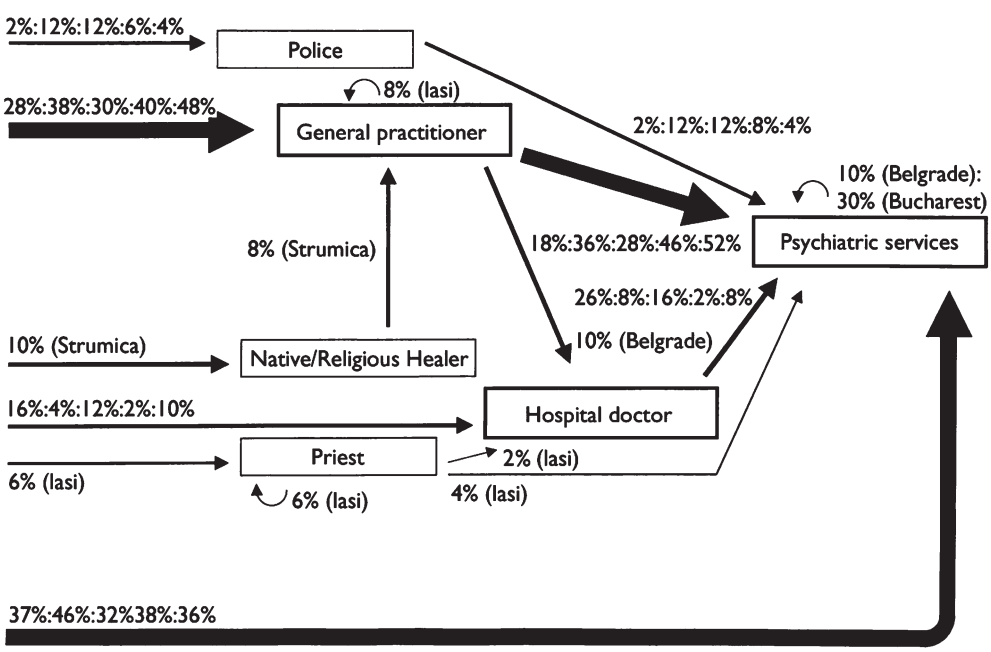


Fig. 7.3.1 Pathways to psychiatric care in Belgrade, Bucharest, Iasi, Strumica and Zagreb. Percentage of those taking each step for each centre respectively, for carers involved in more than 5 per cent of pathways. Steps occurring in only one or two centres are indicated as a single figure followed by the centre name in brackets. (Curved arrows above carer boxes indicate recursive pathways, where patients have gone from one to another of the same type of carer). Reproduced from Gater, R., Jordanova, V., Maric, N. *et al.* (2003). Pathways to psychiatric care in Eastern Europe. *British Journal of Psychiatry*, **186**, 529–35, copyright 2003, The Royal College of Psychiatrists.

depression in South Asian women.[14] Thus, rather than 'ethnic match', the culture of medicine may be the important determinant of recognition of depression in this group and its treatment. Longer delays in first contact with a mental health professional,[10] reduce likelihood of successful drug treatment,[15] and poorer outcomes[16] have been reported in groups of Asian patients. The mismatch between patient experience and culture of medicine appears ubiquitous. In Turkey, psychiatric patients with somatic presentations were shown to have longer delays because they went to see hospital doctors first before being referred to a psychiatrist.[17] A study in Nigeria found that a large proportion of depressed patients initially received another diagnosis because of somatic presentation.[18] However, emphasis on somatic experience on the part of the patient does not preclude the patient's recognition of psychological factors, and addressing culturally shaped experience of illness can help clinicians determine underlying aetiology and understand patients with challenging somatic symptoms.[19]

Beliefs about why the problem has arisen, shape the pathway into care. How important this is depends on how different a culture's models of illness and treatment are from that of the service providers. For example, in a study of help-seeking behaviour of families of patients with schizophrenia in India, most of those who believed in supernatural causation consulted indigenous healers first and those who identified schizophrenia as a medical problem consulted practitioners of modern medicine.[20] Similarly, a study in Ghana found that a perceived supernatural cause of mental health problems was associated with a marked reduction in the likelihood of consulting a mental hospital facility.[21] Surveys among Chinese Americans and Mexican Americans show that[22,23] the association between help-seeking, service and acculturation is mediated, amongst others, by beliefs and explanatory models of psychiatric symptoms. Patient satisfaction is highest if the explanatory model of the patient is matched with that of the service provider.[24] Mental health-care professionals are also prone to differences in explanatory models. For example, a comparison between mental health-care professionals in Saudi Arabia and the UK revealed that the staff in the UK believed in a greater range of possible causes and diagnoses for auditory hallucinations than staff in Saudi Arabia. Differences in belief were associated with different expectations regarding efficacy of possible treatments.[25]

Pathways to care are shaped by public opinion and stigma. Such factors can also contribute to the resources that a society will give, and where and by whom treatment is given. A population survey in Germany showed that the lay public generally held psychotherapy in high esteem and the vast majority of respondents rejected pharmacotherapy for psychological problems. Psychoanalysis was the most popular approach in the western part of Germany but in eastern Germany the preference was for group therapy.[26]

Similarly, because psychiatric practice remains to a large degree opinion-based within Europe, differences exist between national samples of psychiatrists with regard to the diagnosis, aetiology, treatment, and outcome of psychiatric disorders such as schizophrenia. Cultural divergence is especially evident with regard to differential emphasis on psychodynamic and biological approaches,[27] and the level of availability of psychotherapy within the health service in European countries appears to be associated with the dominant therapeutic culture of psychiatrists.[28]

The quality of mental health legislation, and, perhaps more importantly, the degree to which correct implementation is enforced have an effect on care pathways. Increased stigma of psychotic illness and fear of public safety have led to compulsory treatment in the community becoming law in the UK. This could change pathways to and through care.[29] In the Netherlands, a weak and impractical act has been criticized for denying patients with severe mental illness the treatment they need. In Spain, to date no specific mental health act exists. In countries like the United Kingdom and France, involuntary admission is in practice, a largely clinical decision, and therefore more easily to put into effect compared with, for example, The Netherlands and most of Germany where the decision is ultimately made by the judiciary. In the United Kingdom and France, the police may bypass medical referral and take people behaving strangely in a public place directly to a psychiatric hospital. This has been shown to be an important pathway for certain ethnic minority groups, such as African-Caribbeans in the United Kingdom.[30]

Filters on the pathway to care and service use

Filter permeability and service use

The majority of people in psychological distress never present to formal health services. If formal help is sought, the likelihood of receiving such help is influenced by the permeability of a number of filters (see Chapter 7.8). For example, once the decision to seek formal help is taken, actual receipt of help depends on the ability of the professional at the first port of call (usually primary care services) to recognize the presence of mental disorder. The permeability of subsequent filters determines the rate with which patients move from primary care service to specialist mental health outpatient services, and from specialist outpatient to hospital-based psychiatric services and back down the hierarchy.

(a) Cross-ethnic differences

There are important cross-ethnic and cross-national differences in the permeability of filters along the pathway to care. Recognition of symptoms by mental health professionals is an important factor. Recognition of distress is dependent on the way symptoms are elicited. For instance, the recognition by primary care doctors of psychiatric disorder in African-Caribbeans and South Asians in the United Kingdom has been shown to be poor.[31] There are a number of reasons for this including differences between the symptoms expected by doctors and those presented.[12] Despite the fact that they visit their general practitioner more often, women of South Asian origin with depression are less likely to be diagnosed and treated than white British women. Detection of depression depends on the doctor's skill but also whether the patient tells the general practitioner about her worries. Those who believed that a doctor was the right person to deal with depression were found to be more likely to disclose information and more likely to be diagnosed and treated.[12] South Asian women are less likely than white British women to think that a doctor was the right person to deal with depression.[32] A one-year follow-up of the sample of the Epidemiologic Catchment Area Program revealed that African-Americans, Hispanics, and other minorities were much less likely to have consulted with a professional in the specialized mental health care sector than white people. The odds of consultation in African-Americans was less than one-quarter of that in white people even after adjustment for confounders.[33] Similarly,

African-American children and adolescents may also remain under-treated although they may have higher levels of symptomatology.[34] Differences between American ethnic groups are also apparent in populations with identical insurance coverage.[35] Hence, these findings[36] suggest low permeability of filters on the pathway to mental health care. Reasons for this may include that African-Americans are less inclined to seek professional help because of increased tolerance to depressive symptoms, but also because of fear of hospital admission.[37] In comparison to all other ethnic groups, African-Americans make more use of emergency rooms for routine psychiatric care.[38]

Despite the low permeability of the filters on the pathway to care there is an over-representation of African-Caribbeans and African-Americans at the level of hospital-based psychiatric services. Possible mechanisms for this include failure of community services to engage mentally ill African-Caribbean men[39] and bypass of the usual filters by, for example, compulsory admission to hospital with or without police involvement.[40] It has been shown that police involvement and compulsory admission to hospital is strongly associated with the absence of general practitioner involvement.[41] Levels of perceived violence and rates of involuntary admission may be due to stereotyped attitudes of the police and mental health professionals or may be in part due to a higher rate of presentation of psychosis that is superimposed on intact premorbid personalities. It has been suggested that reactive forms of psychotic illness in African-Caribbeans are wrongly labelled as schizophrenia.[42] Higher functioning, less withdrawn patients may be perceived as constituting a higher risk by police and mental health professionals. Another factor is that, despite low rates of recognition by general practitioners, African-Caribbeans are most likely to be referred on to a specialist, followed by white people and then people from south Asia even when socioeconomic class and diagnosis are taken into account.[31]

(b) Cross-national differences

Cross-national differences in filter permeability and service use are difficult to examine. Service organization plays an important role. For example, compared with south Manchester, in the United Kingdom, closer integration between community and inpatient psychiatric services in south Verona, Italy, resulted in a greater permeability of the filter between inpatient and community care, as evidenced by higher hospital admission rates and shorter lengths of stay. Conversely, greater permeability of the general practitioner referral filter in south Manchester resulted in more referrals, and therefore higher treated incidence and prevalence rates of psychiatric disorder.[43] Similarly, compared with the more institution-based system in Groningen, in The Netherlands, the south Verona community-based system provided for a higher degree of continuity of care across services for patients with schizophrenia.[44]

Because of convergence in methodology, fascinating material on cross-national differences in the dynamic balance between psychological distress, need for care, and actual treatment received is now available from large prevalence surveys of representative samples in different countries. Comparative data are available on three large population studies in three countries. In the United States and Ontario (Canada), samples representative of all non-institutionalized individuals aged 18 to 54 years were examined using the Composite International Diagnostic Interview (CIDI) in 1990 and samples of people aged 18 to 64 years in The Netherlands were examined in 1996 (Nemesis Study).[45–47]

The three studies show that help-seeking rates among individuals with a diagnosable disorder vary widely. The contact rate with any type of formal or informal service was lowest in the United States (33.9 per cent) and not far from twice as high in The Netherlands (56.7 per cent) (Table 7.3.1). Among individuals with a diagnosable disorder, ambulatory service use in the general medical sector was much higher in Ontario and The Netherlands than in the United States, especially among those with more severe comorbid

Table 7.3.1 Prevalence of CIDI disorder and service use in three countries

		Service use			
	12-month prevalence	General medical[a]	Ambulatory mental health[b]	Any service	Perceived need for care in non-users of professional services
National Comorbidity Survey					
No disorder[c]	54.0	2.6	2.6	7.6	
One disorder	1S.7	5.6	8.3	18.8	8.4
Two disorders	12.0	10.2	18.6	33.9	
Ontario					
No disorder'	65.6	1.5	1.2	3.3	
One disorder	127	9.7	6.3	17.8	5.4
Two disorders	5.9	24.3	23.5	39.4	
Nemesis					
No disorder[c]	58.8	3.8	1.8	6.1	
One disorder	15.3	15.5	7.8	22.5	3.8
Two disorders	8.1	42.8	30.3	56.7	

[a] Non-psychiatrist physician in any setting or allied health professional in a general medical setting.

[b] Psychiatrist or psychologist in any setting or allied health professional (e.g. nurse, social worker) in a psychiatric or addiction treatment setting.

[c] No lifetime history of any disorder at all.

Weighted data from Kessler *et al.*[45] and Bijl *et al.*[47]

disorders. Individuals in the United States with more severe disorders were less likely to use services in the health-care sector as a whole, but if treatment from self-help and other sources is included, the difference between Ontario and Nemesis on the one hand, and the United States on the other, is attenuated. This suggests that low permeability of the primary care access and primary care referral filters in the United States may lead to increased use of non-professional services to fill the gap.

Table 7.3.1 shows that the higher the contact rates with professional services of individuals with a CIDI diagnosis, the lower the number of individuals who were not using professional services but felt they were in need of such help (level of unmet need) (Table 7.3.1). Such differences in the population level of unmet need are important from the point of view of public health. For example, if 90 per cent in a population of 200 million are non-users of professional services for mental health problems, then the difference between 8.4 per cent (United States) and 3.8 per cent (The Netherlands) in perceived need is a difference of 9.2 million individuals.

The substantial differences in mental health care provided by the general practitioner are likely to have an equally substantive impact on the likelihood of receiving appropriate management. Thus, the proportion of patients with major depression (as defined in DSM-IIIR) in the previous 12 months who received appropriate medication management, defined as a combination of antidepressant medication use and four or more visits to any health-care provider within the previous 12 months, was much higher in Ontario (14.9 per cent) than in the United States (7.3 per cent). This difference was especially marked for the lowest income groups in the two countries. Individuals in the lowest income groups in the United States were found to be 7.5 times less likely to make contact with either general or specialty health-care providers than their peers in Ontario. For the highest income groups, however, contact rates differed only by a factor 2.1.[48] These data suggest that economic barriers play an important role in determining the permeability of the filters on the pathway to care. In a United Kingdom national survey, 16 per cent of patients with a depressive episode in the past week according to the Revised Clinical Interview Schedule were current users of antidepressant medication.[49]

Because the majority of the population does not have a mental disorder, even a small degree of service use by this large segment of the population will take up a considerable part of the total capacity of mental health services. Thus, Katz *et al.*[50] noted that because of the relatively high rates of perceived need for care and help-seeking among individuals without a CIDI diagnosis in the United States, total mental health outpatient service use was higher in that country than in Ontario. Although diagnosis is only an imperfect indicator of need for care, the results nevertheless suggest that the mismatch between need and care in the population is greater in the United States than in Ontario.

A long-standing debate exists whether and how financing of mental health care can be used to maximize the fit between need and care in the population. A frequently expressed concern is that universal coverage will lead to an increase of people with little need using services of unproven value. The opposite argument, however, is that limitations in coverage will result in service use that is poorly matched to need. Although it is thought that differences in type of insurance system have an impact on demand and utilization of mental health services,[51] systematic comparisons between countries have been lacking. The systems of coverage in the United States, Canada, and The Netherlands are different in many respects. In Ontario, universal and relatively comprehensive coverage for mental health services exists, with no or minimal limits on inpatient stays or outpatient visits for mental health services, and minimal patient cost sharing. In The Netherlands, almost all mental health care is covered under the Exceptional Medical Expenses Act, and is available to the entire population. A comprehensive range of public services exists, with few supply-side controls. In the United States, at least 16 per cent of the population is uninsured, and even for the insured mental health coverage is increasingly limited. Although the public health system provides mental health care at little or no cost to the poor and the uninsured, supply-side controls severely and increasingly limit access. Therefore, the results of the comparisons between the three countries do not support the frequently expressed reservation that expansion of insurance coverage for mental health disorders results in an increase in unnecessary use of services. Of the three countries considered, those with broad mental health coverage actually treated a similar number or more people with severe mental illness, but less people who never had a history of mental illness.

Treatment and response

Culture has an important influence on the type of service received. There are reports that African-Caribbean and African-American patients receive antipsychotic medications in a higher prescribed dose, and more frequent use of injectable preparations is made.[30,52–57] Asian patients have been reported to receive lower doses.[58,59] Some ethnic minority groups may receive less information about side-effects,[60] which may result in less vigilance with regard to onset of problems such as tardive dyskinesia. In the United States and the United Kingdom, ethnic minority patients are less likely to receive psychological treatment.[61,62] In the United States, African-American children have substantially lower rates of receiving methylphenidate.[63,64] Such differences may be related to differences in explanatory models of African-American parents, and differences in the rate with which African-American parents receive appropriate information about attention deficit–hyperactivity disorder from the doctor.[65]

There are several important considerations with regard to outcome in relation to cultural variables.[66] The first is that services offered may not be equally efficacious for different groups of people. For example, Chinese, Japanese, Filipino, Korean, and Southeast Asian Americans who were treated in the same setting in Los Angeles County showed different outcomes. Filipinos were under-represented in the system, whereas Southeast Asians were over-represented and had higher rates of service utilization. Despite this, Southeast Asians showed less improvement than the other groups, even after controlling for diagnosis and initial level of functioning.[67] The second is that treatment uptake may differ between groups.[33] In the United Kingdom and the United States, uptake of treatment with antipsychotic medication may be higher in white patients, though the overall influence of ethnicity remains small.[68] The third consideration is that the expectations about the desired endpoint of treatment may not be the same for different groups of people. For example, the expectations of British Asian and white people relating to the process and outcome of a psychological intervention were shown to be different in one study.[69]

Outcome may improve if therapists receive information about their clients' cultural background and expectations before treatment.[70] Perhaps the most important consideration is that outcome is a multidimensional concept defying summary statements. For example, clinical outcome in terms of usual symptom severity and risk of self-harm may be better in African-Caribbean patients with psychosis as compared with white people, yet risk of imprisonment and compulsory admission may be greater,[71] as may be the frequency of relapse[72] and the rate of dissatisfaction with services.[73]

However, it is interesting to note that a recent pan-European study has concluded that despite the fact that differences in health-care systems may affect service provision and cost, the impact of such differences on outcome may be less marked. More work needs to be undertaken on cultural differences in care pathways and treatment to see if they change outcomes.[74] Moreover, at an individual level, recent work has failed to show differences in the duration of untreated psychosis between African Caribbeans and whites in the UK,[75] even though they have different pathways to care at first admission indicating that the impact of cultural or ethnic differences in pathways can be difficult to predict.

Pathways out of care

Pathways out of care are complex and have multiple influences. An important influence is the pathway into care because the relationship set up with services at first contact determines, in part, the trajectory of the clinical career of a patient. Moreover, in specific circumstances, such as where the criminal justice system is involved, certain responsibilities may be placed on services which constrain the ability to discharge patients. There has been little research which aims to understand ethnic differences in pathways out of care.[76] This reflects the fact that the focus of studies to date has been to document any inequalities in access to care.

Pathways out of statutory sector care are influenced by the effectiveness of treatment strategies, the illness models and treatment preferences of patients, the structure and funding of clinical services, the availability of and access to nonstatutory sector support and also the socio-cultural context in which people live.

Effectiveness is a combination of the efficacy of a treatment and the real world context which influences outcomes. Though there may be differences in the efficacy of drugs by racial group, differences in compliance are more likely to be important,[77] and there are well documented differences in adherence to treatment programs which may lead to differences in rates of patients leaving care.[78]

Adherence has been linked to differences in illness models but in psychosis the level of discord between patient and treatment service illness models are not consistently associated with differences in outcome.[79] This may reflect the fact that preferred alternative treatment outside statutory mental health services are not available or that services are more coercive with patients who do not fully subscribe to their model.

The structure and funding of clinical services is important. Filters out of care can be porous—an example of this is Ontario, where rehabilitation is mainly offered by local community based organizations which may have better links with the community and so make transition out of care easier. They also have limited responsibility for their clients compared to the UK and may be less risk averse. Risk averse service cultures can lead to delay in discharge from care.[80] Moreover, putting the onus on risk can mean that ethnic groups who are considered more risky have particular difficulty in getting out of care.

Prevailing law also shapes pathways out of care. Such laws may be applied differently to some ethnic groups. For instance, in New Zealand, the Maori population is more likely to be detained on community treatment orders thus delaying movement out of statutory service care.[81]

The availability of nonprofessional support and the structure of some cultural groups may affect pathways. Those of South Asian origin with severe mental illness spend less time in hospital and outpatient care. Family support is cited as the reason for this. Socio-cultural mores are important but also geographic concentration; dispersed refugee and asylum groups with limited access to community support may stay longer in services than would be expected from their symptoms. But socially cohesive groups can be a double-edged sword. One study has demonstrated higher rates of re-admission in areas with high social capital which arguably could be a result of high levels of health norm policing and low levels of tolerance for deviance.[82]

Socio-economic factors also need to be kept in mind when considering the capacity of communities to offer care. The decreased resources available to low income groups will be important in determining the level of burden they will be able to accept.

Further information

World Health Organization: up to date information on the health with a good archive, research tools and data as well as a section which summarizes health systems organization and funding for each of the countries in the world. http://www.who.int/en/

World Association of Cultural Psychiatry: free newsletter and journal on cross cultural psychiatry issues at: http://www.waculturalpsychiatry.org/

National Institute for Clinical Excellence in the UK has developed a pathways approach to improving mental health services for people with schizophrenia backed by algorithms and research evidence: http://www.schizophreniaguidelines.co.uk/nice_implementation/pathways_to_care.php.

References

1. Rogler, L.H. and Cortes, D.E. (1993). Help-seeking pathways: a unifying concept in mental health care. *American Journal of Psychiatry*, **150**, 554–61.
2. Gater, R., de Almeida e Sousa, B., Barrientos, G., *et al.* (1991). The pathways to psychiatric care: a cross-cultural study. *Psychological Medicine*, **21**, 761–74.
3. Üstün, T.B. and Sartorius, N. (eds.) (1995). *Mental illness in genera health care*. Wiley, New York.
4. Gater, R., Jordanova, V., Maric, N., *et al.* (2003). Pathways to psychiatric care in eastern Europe. *British Journal of Psychiatry*, **186**, 529–35.
5. Bhui, K., Stansfeld, S., Hull, S., *et al.*(2003). Ethnic variations in pathways to and use of specialist mental health services in the UK. Systematic review. *British Journal of Psychiatry*, **182**, 105–16.
6. Wheeler, A., Robinson, E., Robinson, G., *et al.* (2005). Admissions to acute psychiatric inpatient services in Auckland, New Zealand: a demographic and diagnostic review. *NZ med J*, **118** (1226): u1752.
7. Malla, A. and Norman, R.M.G. (1988). Involuntary admissions in a Canadian province: the influence of geographic and population factors. *Social Psychiatry and Psychiatric Epidemiology*, **23**, 247–51.
8. Engberg, M. (1991). Involuntary commitment in Greenland, the Faroe Islands and Denmark. *Acta Psychiatrica Scandinavica*, **84**, 353–6.

9. Lin, T.Y., Tardiff, K., Donetz, G., *et al.* (1978). Ethnicity and patterns of help-seeking. *Culture in Medicine and Psychiatry*, **2**, 3–13.
10. Lin, K.M., Inui, T.S., Kleinman, A.M., *et al.* (1982). Sociocultural determinants of the help-seeking behavior of patients with mental illness. *Journal of Nervous and Mental Disorders*, **170**, 78–85.
11. King, G. (1996). Institutional racism and the medical health complex: a conceptual analysis. *Ethnicity and Disease*, **6**, 30–46.
12. Kleinman, A. (1991). *Rethinking psychiatry*. Free Press, New York.
13. Weiss, M.G., Raguram, R., and Channabasavanna, S.M., *et al.* (1995). Cultural dimensions of psychiatric diagnosis. A comparison of DSM-IIIR and illness explanatory models in south India. *British Journal of Psychiatry*, **166**, 353–9. Erratum: *British Journal of Psychiatry*, **167**, 119 (1995).
14. Jacob, K.S., Bhugra, D., Lloyd, K.R., *et al.* (1998). Common mental disorders, explanatory models and consultation behaviour among Indian women living in the UK. *Journal of the Royal Society of Medicine*, **91**, 66–71.
15. Cornwell, J. (1998). Do general practitioners prescribe antidepressants differently for South Asian patients? *Family Practice*, **15**, S16–18.
16. Ying, Y.-W. and Hu, L.-T. (1994). Public outpatient mental health services: use and outcome among Asian Americans. *American Journal of Orthopsychiatry*, **64**, 448–55.
17. Kilic, C., Rezaki, M., Üstün, T.B., *et al.* (1994). Pathways to psychiatric care in Ankara. *Social Psychiatry and Psychiatric Epidemiology*, **29**, 131–6.
18. Makanjuola, J.D. and Olaifa, E.A. (1987). Masked depression in Nigerians treated at the Neuro–Psychiatric Hospital Aro, Abeokuta. *Acta Psychiatrica Scandinavica*, **76**, 480–5.
19. Handelman, L. and Yeo, G. (1996). Using explanatory models to understand chronic symptoms of Cambodian refugees. *Family Medicine*, **28**, 271–6.
20. Banerjee, G. and Roy, S. (1998). Determinants of help-seeking behaviour of families of schizophrenic patients attending a teaching hospital in India: an indigenous explanatory model. *International Journal of Social Psychiatry*, **44**, 199–214.
21. Fosu, G.B. (1995). Women's orientation toward help-seeking for mental disorders. *Social Science and Medicine*, **40**, 1029–40.
22. Wells, K.B., Hough, R.L., Golding, J.M., *et al.* (1987). Which Mexican-Americans underutilize health services? *American Journal of Psychiatry*, **144**, 918–22.
23. Ying, Y.-W. and Miller, L.S. (1992). Help-seeking behavior and attitude of Chinese Americans regarding psychological problems. *American Journal of Community Psychology*, **20**, 549–56.
24. Callan, A. and Littlewood, R. (1998). Patient satisfaction: ethnic origin or explanatory model? *International Journal of Social Psychiatry*, **44**, 1–11.
25. Wahass, S. and Kent, G. (1997). A cross-cultural study of the attitudes of mental health professionals towards auditory hallucinations. *International Journal of Social Psychiatry*, **43**, 184–92.
26. Angermeyer, M.C. and Matschinger, H. (1996). Public attitude towards psychiatric treatment. *Acta Psychiatrica Scandinavica*, **94**, 326–36.
27. van Os, J., Galdos, P., Lewis, G., *et al.* (1993). Schizophrenia sans frontieres: concepts of schizophrenia among French and British psychiatrists. *British Medical Journal*, **307**, 489–92.
28. van Os, J. and Neeleman, J. (1994). Caring for mentally ill people. *British Medical Journal*, **309**, 1218–21.
29. Department of Health (2007) regulatory impact assessment of mental health bill available at dh.gov.uk/prod_consum_dh
30. Dunn, J. and Fahy, T.A. (1990). Police admissions to a psychiatric hospital. Demographic and clinical differences between ethnic groups. *British Journal of Psychiatry*, **156**, 373–8.
31. Rawaf, S. and Bahl, V. (1998). *Assessing health needs of people from minority ethnic groups*. Royal College of Physicians of London.
32. Lloyd, K.R., Jacob, K.S., Patel, V., *et al.* (1998). The development of the Short Explanatory Model Interview (SEMI) and its use among primary-care attenders with common mental disorders. *Psychological Medicine*, **28**, 1231–7.
33. Gallo, J.J., Marino, S., Ford, D., *et al.* (1995). Filters on the pathway to mental health care. II. Sociodemographic factors. *Psychological Medicine*, **25**, 1149–60.
34. Cuffe, S.P., Waller, J.L., Cuccaro, M.L., *et al.* (1995). Race and gender differences in the treatment of psychiatric disorders in young adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**, 1536–43.
35. Scheffler, R.M. and Miller, A.B. (1989). Demand analysis of mental health service use among ethnic subpopulations. *Inquiry*, **26**, 202–15.
36. Brown, D.R., Ahmed, F., Gary, L.E., *et al.* (1995). Major depression in a community sample of African Americans. *American Journal of Psychiatry*, **152**, 373–8.
37. Sussman, L.K., Robins, L.N., Earls, F., *et al.* (1987). Treatment-seeking for depression by black and white Americans. *Social Science and Medicine*, **24**, 187–96.
38. Neighbors, H.W. (1986). Ambulatory medical care among adult black Americans: the hospital emergency room. *Journal of the National Medical Association*, **78**, 275–82.
39. Bhui, K., Brown, P., Hardie, T., *et al.* (1998). African-Caribbean men remanded to Brixton Prison—psychiatric and forensic characteristics and outcome of final court appearance. *British Journal of Psychiatry*, **172**, 337–44.
40. Morgan, C., Mallett, R., Hutchinson, G., *et al.* (2005). On behalf of the ÆSOP Study Group Pathways to care and ethnicity II; Source of referral and help seeking. A Report From the ÆSOP (Aetiology and Ethnicity in Schizophrenia and Other Psychoses) Study. *British Journal of Psychiatry* ,**186**, 290–6.
41. Cole, E., Leavey, G., King, M., *et al.* (1995). Pathways to care for patients with a fi rst episode of psychosis. A comparison of ethnic groups. *British Journal of Psychiatry*, **167**, 770–6.
42. Littlewood, R. and Lipsedge, M. (1978). Migration, ethnicity and diagnosis. *Psychiatric Clinics of Basel*, **11**, 15–22.
43. Amaddeo, F., Gater, R., Goldberg, D., *et al.* (1995). Affective and neurotic disorders in community-based services: a comparative study in south Verona and south Manchester. *Acta Psychiatrica Scandinavica*, **91**, 386–95.
44. Sytema, S., Micciolo, R., Tansella, M., *et al.* (1997). Continuity of care for patients with schizophrenia and related disorders: a comparative south Verona and Groningen case-register study. *Psychological Medicine*, **27**, 1355–62.
45. Kessler, R.C., Frank, R.G., Edlund, M., *et al.* (1997). Differences in the use of psychiatric outpatient services between the United States and Ontario. *New England Journal of Medicine*, **336**, 551–7.
46. Katz, S.J., Kessler, R.C., Frank, R.G., *et al.* (1997). The use of outpatient mental health services in the United States and Ontario: the impact of mental morbidity and perceived need for care. *American Journal of Public Health*, **87**, 1136–43.
47. Bijl, R.V., Ravelli, A., Van Zessen, G., *et al.* (1998). Prevalence of psychiatric disorder in the general population: results from The Netherlands mental health survey and incidence study. *Social Psychiatry and Psychiatric Epidemiology*, **33**, 587–95.
48. Katz, S.J., Kessler, R.C., Lin, E., *et al.* (1998). Medication management of depression in the United States and Ontario. *Journal of General Internal Medicine*, **13**, 77–85.
49. Meltzer, H., Gill, B., Petticrew, M., *et al.* (1995). *OPCS surveys of psychiatric morbidity. Report 1: the prevalence of psychiatric morbidity among adults aged 16–64 living in private households in Great Britain*. HMSO, London.
50. Katz, S.J., Kessler, R.C., Frank, R.G., *et al.* (1997). The use of outpatient mental health services in the United States and Ontario: the impact of mental morbidity and perceived need for care. *American Journal of Public Health*, **87**, 1136–43.

51. Frank, R.G. and McGuire, T.G. (1986). A review of studies of the impact of insurance on the demand and utilization of specialty mental health services. *Health Services Research*, **21**, 241–65.
52. Citrome, L., Levine, J., Allingham, B., *et al.* (1996). Utilization of depot neuroleptic medication in psychiatric inpatients. *Psychopharmacology Bulletin*, **32**, 321–6.
53. Segal, S.P., Bola, J.R., Watson, M.A., *et al.* (1996). Race, quality of care, and antipsychotic prescribing practices in psychiatric emergency services. *Psychiatric Services*, **47**, 282–6.
54. Lawson, W.B. (1996). Clinical issues in the pharmacotherapy of African-Americans. *Psychopharmacology Bulletin*, **32**, 275–81.
55. Price, N., Glazer, W., Morgenstern, H., *et al.* (1985). Demographic predictors of the use of injectable versus oral antipsychotic medications in outpatients. *American Journal of Psychiatry*, **142**, 1491–2.
56. Shubsachs, A.P., Huws, R.W., Close, A.A., *et al.* (1995). Male Afro-Caribbean patients admitted to Rampton Hospital between 1977 and 1986—a control study. *Medicine, Science and Law*, **35**, 336–46.
57. Strakowski, S.M., Shelton, R.C., Kolbrener, M.L., *et al.* (1993). The effects of race and comorbidity on clinical diagnosis in patients with psychosis. *Journal of Clinical Psychiatry*, **54**, 96–102.
58. Rosenblat, R. and Tang, S.W. (1987). Do oriental psychiatric patients receive different dosages of psychotropic medication when compared with occidentals. *Canadian Journal of Psychiatry*, **32**, 270–4.
59. Bond, W.S. (1991). Ethnicity and psychotropic drugs. *Clinical Pharmacology*, **10**, 467–70.
60. Benson, P.R. (1984). Drug information disclosed to patients prescribed antipsychotic medication. *Journal of Nervous and Mental Disease*, **172**, 642–53.
61. Olfson, M. and Pincus, H.A. (1994). Outpatient psychotherapy in the United States. II. Patterns of utilization. *American Journal of Psychiatry*, **151**, 1289–94.
62. Littlewood, R. and Lipsedge, M. (1997). *Aliens and alienists*. Routledge, New York.
63. Zito, J. M., Safer, D. J., dos Reis, S., *et al.* (1997). Methylphenidate patterns among Medicaid youths. *Psychopharmacology Bulletin*, **33**, 143–7.
64. Zito, J. M., Safer, D. J., dos Reis, S., *et al.* (1998). Racial disparity in psychotropic medications prescribed for youths with Medicaid insurance in Maryland. *Journal of the American Academy of Child and Adolescent Psychiatry*, **37**, 179–84.
65. Bussing, R., Schoenberg, N.E., Perwien, A.R., *et al.* (1998). Knowledge and information about ADHD: evidence of cultural differences among African-American and white parents. *Social Science and Medicine*, **46**, 919–28.
66. McKenzie, K. and Murray, R.M. (1998). *Risk factors for mental illness in African-Caribbeans.* Department of Health Conference on Culture, Ethnicity and Mental Health. Department of Health, London.
67. Ying, Y.W. and Hu, L.T. (1994). Public outpatient mental health services, use and outcome among Asian Americans. *American Journal of Orthopsychiatry*, **64**, 448–55.
68. Bebbington, P.E. (1995). The content and context of compliance. *International Clinics in Psychopharmacology*, **5**, 41–50.
69. Balabil, S. and Dolan, B. (1992). A cross-cultural evaluation of expectations about psychological counselling. *British Journal of Medical Psychology*, **65**, 305–8.
70. Yamamoto, J., Acosta, F. X., Evans, L.A., *et al.* (1984). Orienting therapists about patients' needs to increase patient satisfaction. *American Journal of Psychiatry*, **141**, 274–7.
71. McKenzie, K., van Os, J., Fahy, T., *et al.* (1995). Psychosis with good prognosis in Afro-Caribbean people now living in the United Kingdom. *British Medical Journal*, **311**, 1325–8.
72. Birchwood, M., Cochrane, R., Macmillan, F., *et al.* (1992). The influence of ethnicity and family structure on relapse in first-episode schizophrenia. A comparison of Asian, Afro-Caribbean, and white patients. *British Journal of Psychiatry*, **161**, 783–90.
73. Parkman, S., Davies, S., Leese, M., *et al.* (1997). Ethnic differences in satisfaction with mental health services among representative people with psychosis in south London: PRiSM Study 4. *British Journal of Psychiatry*, **171**, 260–4.
74. Becker, T., Kilian, R. (2006). Psychiatric services for people with severe mental illness across western Europe: what can be generalized from current knowledge about differences in provision, cost and outcomes of mental health care *Acta Psychiatr Scan,* Suppl 429, 9–16
75. Morgan, C., Fearon, P., Hutchinson, G., AESOP Study Group., *et al.* (2006) Duration of untreated psychosis and ethnicity in the AESOP fi rst-onset psychosis study. *Psychological Medicine,* **36**(2), 239–47
76. EPIC group (2007). ww.wolfson.qmul.ac.uk/psychiatry/epic/docs/SRper cent20Keyper cent20Summary.pdf accessed
77. Lin, K.M., Anderson, D., Poland, R.E., *et al.* (1995). Ethnicity and psychopharmacology. Bridging the gap. *Psychiatr Clin North Am*. **18**(3), 635–47.
78. Sue, S., Fujino, D. C., Hu, L. T., *et al.* (1991). Community mental health services for ethnic minority groups: a test of the cultural responsiveness hypothesis. *J Consult Clin Psychol*, **59**(4), 533–40.
79. Rosemarie McCabe, PhD and Stefan Priebe, MD. (2004). Explanatory models of illness in schizophrenia: comparison of four ethnic groups *The British Journal of Psychiatry,* **185**, 25–30
80. Lelliott, P. and Audini, B. (2003). Trends in the use of Part II of the Mental Health Act 1983 in seven English local authority areas. Br J Psychiatry, **182**, 68–70.
81. Gibbs, A., Dawson, J., Forsyth, H., (2004). Maori experience of community treatment orders in Otago, New Zealand. *Aust N Z J Psychiatry*, **38**(10), 830–5s
82. McKenzie, K. (2000). Neighborhood, safety and mental health outcomes. Postingnumber 28, 2000. Social Capital Lets Talk. Socialcapital@tome. worldbank.org

7.4

Primary prevention of mental disorders

J. M. Bertolote

Despite the demonstration of the possibility of preventing some forms of mental disorders, many mental health professionals continue to underestimate the possibilities of primary prevention in their field. This is due to:

1 a lack of clear concepts when referring to this issue;

2 the fact that the effective prevention of mental and neurological disorders often falls outside the usual remit of mental health professionals (in many cases it falls outside the health sector altogether).

These two factors are discussed below, in addition to an indication of actions which effectively prevent some forms of mental disorders.

Prevention

In the late 1950s, Leavell and Clark[1] proposed a three-level concept of prevention (primary, secondary, and tertiary), covering almost all medical actions. Their innovative approach must be understood in relation to what they also called the horizon of the natural history of the disease process: under natural circumstances, a disease will proceed from a prepathological period through its early stages, evolving either to partial or full recovery (or cure), or to death; in the case of partial recovery, there may be chronification or sequelae. Prevention, in this sense, refers not only to the appearance of the disease but also to any further worsening or complication of it once it has appeared.

Primary prevention

The primary prevention level covers what is otherwise referred to as both health promotion and specific protection, and is best exemplified by, for example, adequate nutrition and immunization against specific diseases by vaccines. Whereas, in this example, adequate nutrition is totally non-specific (it contributes to enhancing the overall resistance to several diseases without conferring any specific protection against any), vaccination is highly specific in relation to a single condition. Rational-specific protection is fully dependent on a reasonable knowledge of the aetiology of the disease (or, at least, its mode of transmission) in order to be effective.

Secondary prevention

The secondary prevention level refers to early detection and treatment of diseases. Usually the bulk of the medical activity, its main preventive goal is to avoid chronicity and the establishment of irreversible sequelae. It is dealt with more specifically in Part 6 of this book.

Tertiary prevention

The tertiary prevention level largely corresponds to rehabilitation. It enters into operation once the disease process has been established and aims at reducing as much as possible damages caused by the disease process, preserving intact functions, and restoring and/or compensating impaired functions, disabilities, and handicaps.

On one hand, this conceptual model was highly instrumental in providing an impetus towards preventive activities in the medical field as a whole but, on the other hand, it so popularized the term prevention that it almost lost its powerful message. Therefore, it is important to retain the idea of primary prevention as a synonym of specific protection, referring to methods designed to avoid the occurrence of a specific disorder or groups of disorders. It comprises those measures applicable to a particular disease or group of diseases in order to intercept their causes before they affect people, and should be differentiated not only from treatment and rehabilitation, but also from mental health promotion.

The main obstacle for the prevention of many mental disorders is the limited knowledge about their aetiology. Admittedly, there are very promising and exciting hypotheses concerning the causes of three mental disorders which represent the greatest burden, namely, depression, schizophrenia, and dementia. They are, nevertheless, nothing more than hypotheses. The most successful examples of prevention of diseases refer to those whose aetiology (cause and/or mode of transmission) is relatively well-known. There are historical examples of the prevention of some conditions based on false assumptions about or without a good knowledge of their aetiology (for example, the eradication of malaria in ancient Rome, and the control of the London cholera epidemics by John Snow in the nineteenth century). However, it does not seem appropriate for health professionals and scientists to base their actions on chance or false assumptions, even though the result might be opportune to the population.

As implied by the need to intercept causes of a particular disease or groups of diseases (with a common cause), the concept of prevention calls for a high degree of specificity concerning the target condition or conditions. In the medical field, it led to successful programmes for the prevention of, for example, diarrhoeal diseases (such as typhoid), hypertension, coronary heart disease, breast cancer, and unwanted pregnancies, rather than of infectious diseases, cardiovascular diseases, cancer, or obstetrical problems. Unfortunately, in the mental health field there has not been a great concern with the specification of the target condition, and the prevention of 'mental disorders' (as a whole) became a label soon associated with failure and disinterest.

Mental disorders

What is understood as 'mental disorders' comprises a variety of quite diverse clinical conditions in terms of aetiology, symptomatology, clinical course, prognosis, and response to treatment. Therefore, whenever the prevention of mental disorders is referred to, an effort must be made to obtain some precision.

From a nosological point of view, most of the mental disorders are conceptually at a syndromal level; depression, schizophrenia, and dementia are appropriate examples. In this respect, dementia is one step ahead of the other two, in so far as vascular dementia is now clearly differentiated from Alzheimer's disease, with important implications for prevention.

Therefore a strategic shift is necessary in order to obtain greater efficiency in the successful prevention of some mental disorders. The first step is for an effort to be as specific as possible in relation to the target condition: for instance Down syndrome or phenylketonuria instead of intellectual disorder, foetal alcohol syndrome, delirium tremens instead of alcoholism, and vascular dementia and dementia following brain injury instead of dementia in general.

The second step applies to those conditions which cannot be meaningfully broken down into more specific conditions, such as schizophrenia or depression. In these cases, the target is displaced from the appearance of the conditions towards future relapses, once a first episode has occurred; this conveniently applies to schizophrenia, depression, and dependence on alcohol and other drugs.

Finally, there are some violent behaviours, such as suicide, parasuicide, and violence against others, the control (and prevention) of which are largely expected by society to come from the field of mental health. They do not characterize a mental disorder in particular, but are frequently associated with one or more of them. Their prevention, therefore, requires specifically dedicated interventions.

With this wide range of issues considered as mental disorders, it becomes clear that the coverage of their prevention goes well beyond the limits of this chapter. A detailed conceptual approach to the prevention of mental and psychosocial disorders can be found in a recent publication of the World Health Organization (**WHO**).[2]

Prevention of mental disorders

From a practical point of view there are three groups of conditions for which efficient preventive action has been documented.

1 Mental disorders with known aetiology: this mostly includes those disorders demonstrated to have an organic basis, ranging from the 'historical' general paresis and dementing disorders (e.g. vascular dementia, pellagra, and dementias associated with infectious and parasitic diseases such as malaria and HIV infection) to several forms of intellectual disorder (Down syndrome, foetal alcohol syndrome, phenylketonuria, and intellectual disorder due to iodine deficiency).

2 Mental disorders without a well-established aetiology but with a relatively predictable course: these are chronic disorders with a recurrent relapsing fluctuating pattern, such as schizophrenia, mood disorders (unipolar and bipolar), and alcohol dependence syndrome.

3 Psychosocial problems strongly associated with mental disorders: these range from violence (domestic and other) to suicide and staff burnout.

Mental disorders with known aetiology

(a) Infectious diseases

Prevention of this group of disorders has by far yielded the greatest success. The demonstration in 1911 by Noguchi and Moore of the brain infection by *Treponema pallidum* as the cause of general paresis[3] opened the way in 1917 to its treatment by malaria therapy, and later to its prevention with penicillin; this is now a landmark in the history of medicine. The discovery of the aetiology of pellagra also led to its prevention and control, leading to the prevention of one type of dementia associated with alcoholism and avitaminosis.

These two once very frequent diseases have almost completely disappeared and there are many experienced psychiatrists who never come across a single case of either; with them also disappeared the history of their successful control. Although the same success has not yet been achieved in relation to vascular dementia, the control of hypertension and atherosclerosis (e.g. through the reduction of salt and fat intake) can significantly reduce brain damage and ensuing dementia (vascular or multi-infarct dementia).

In some developing countries, meningitis and malaria (and, to a lesser extent, inadequately treated epilepsy) are important causes of permanent brain damage which can also lead to dementing disorders. The environmental control of malaria and other brain infections, of which bacterial meningitis is the most important, and their early and prompt treatment can reduce the impact of the infection on the brain and prevent these forms of dementia (or intellectual disorder, depending on the age of onset).

More recently, it has been demonstrated that in some people infected with HIV, the initial manifestations of AIDS are accompanied by some forms of mental disorder, such as mood disorders or dementia.[4] The prevention of these forms of mental disorders follow the same measures as for the prevention of AIDS in general. However, it is not yet certain if the newer combined treatments (bi- and tritherapy) can alter the course of AIDS when brain damage due to HIV has been confirmed.

(b) Intellectual disorder (mental retardation)

Up to 15 per cent of cases of intellectual disorder could be prevented by dealing with the causes that lead to it. A recent WHO publication[2] has set detailed guidelines for the prevention of some forms of this condition, namely, Down syndrome, foetal alcohol syndrome, phenylketonuria, and iodine deficiency syndrome.

These preventive actions are both efficient and affordable even in very poor regions of the world.

- *Down syndrome*—The primary prevention of Down syndrome can be successfully achieved through the control of the age at which women become pregnant: ideally, the age range during which the risk is minimal is between 16 and 35 years, after which the risk increases almost exponentially, as shown in Fig. 7.4.1. Amniocentesis is a procedure that can be very useful for the *in utero* diagnosis of Down syndrome (as well as of other problems and malformations). Where it is culturally and morally acceptable, and legally permitted, a therapeutic abortion is viewed by some as another primary prevention measure.
- *Iodine deficiency*—The world population at risk of intellectual disability due to iodine deficiency is approximately 1 billion and it still occurs in large numbers in some regions of the globe.[6] However, it can be very efficiently and cheaply prevented through the addition of iodine to salt, milk, flour, or water, or, in special situations, through injections of an oily solution containing iodine.[7]
- *Phenylketonuria*—Intellectual disorder due to phenylketonuria can also be successfully prevented through the early identification of children at risk who then receive a phenylalanine-free diet throughout their lives.[8]
- *Foetal alcohol syndrome* (FAS)—Intellectual disorder and malformations seen in FAS syndrome can be prevented if women stay away from alcohol during pregnancy, more particularly during the first trimester, or at least keep their alcohol intake below the dangerous limit of 15 g of ethanol per day.[9]

Table 7.4.1 summarizes actions which can effectively prevent some forms of intellectual disorder. Prevention of intellectual disorder is discussed further in Chapter 10.3.

Table 7.4.1 Action to prevent mental retardation

Condition	Preventive action
Iodine deficiency disorders	Iodize salt or water supply Treat individuals at risk with iodized oil or Lugol's solution
Down syndrome	Discourage pregnancies in women over the age of 35 years If appropriate, provide amniocentesis to women over the age of 35 years
Foetal alcohol syndrome	Use simple screening test to identify women at risk Discourage women from drinking alcohol during pregnancy Alert women that drinking around the time of conception increases the risk to the child
Phenylketonuria	Screen all newborn babies for phenylketonuria Treat with a special low-phenylalanine diet Discourage pregnancies in women with phenylketonuria

Mental disorders without a well-established aetiology but with a relatively predictable course

In this group of disorders, the target for prevention is not the disorder itself, whose aetiology is not clearly established, but the occurrence of further episodes of the disorder, since these are chronic recurrent disorders. Three good examples of this are mood disorders, schizophrenia, and alcohol dependence syndrome.

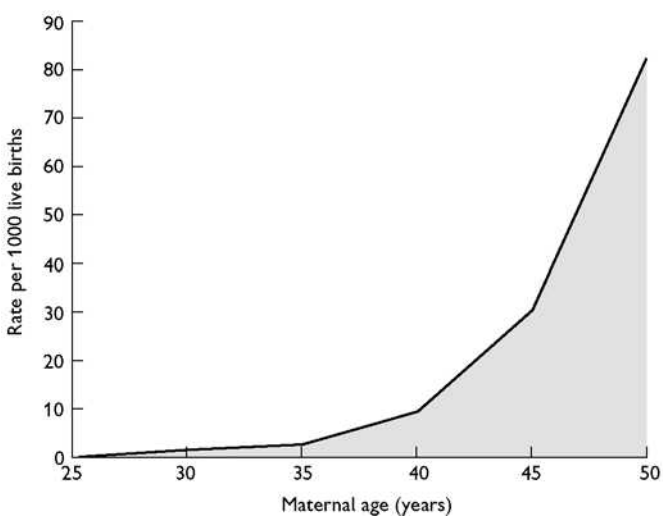


Fig. 7.4.1 Estimated risk of Down syndrome by related age. (Data from Gottesman.[5] Taken from the contribution of genetic factors to the common psychopathologies © World Health Organization, www.who.int)

(a) Mood disorders

In addition to the use of antidepressants and neuroleptics to treat episodes of depression and mania, respectively, it has been demonstrated that the appropriate use of lithium salts can prevent the reappearance of new episodes of disease, or at least increase the disease-free periods and reduce both their duration and severity. The use of lithium salts, also called psychoprophylactics, is now a standard procedure for the treatment and management of mood disorders and this use can be considered as a form of prevention of mood disorders (see Chapter 4.5.8).

(b) Schizophrenia

Although no evidence of a definite cause of schizophrenia is yet available, it can be reasonably controlled through the use of psychopharmacotherapy and psychosocial interventions. Relapse rates (measured by number of hospital readmissions or days in hospital) decrease significantly when persons with schizophrenia adhere to some specific pharmacological regimens and are exposed, together with their relatives, to psychoeducational programmes.[10] These usually include some social skills training for the patient and information about the disease and the management of expressed emotions for the relatives. Unfortunately, schizophrenia is a very long-lasting condition, even lifelong, and most studies on the combination of those two approaches have not yet gone beyond a follow-up of 48 months, thus limiting the full appreciation of its long-term efficiency.

(c) Alcoholism

This term refers to both the alcohol dependence syndrome and other forms of problematic use of alcohol. For people in either category, whenever total abstinence is an unachievable or undesirable goal, the intake reduction, particularly when in connection with risk or dangerous situations, may become the target for prevention. There are several therapeutic techniques usually revolving around brief interventions that are useful to help people significantly to reduce their alcohol and their exposure to problems

associated with it (harm reduction). One should neither forget nor minimize the positive impact of self-help groups (such as Alcoholics Anonymous) in helping some people to achieve sobriety.

Psychosocial problems strongly associated with mental disorders

(a) Violent behaviour

In this category, the immediate target condition is not strictly a mental disorder. However, its close connections with some forms of mental disorder or symptoms make their prevention of immediate concern in the mental health field. Domestic violence (spouse and child beating) is strongly associated with substance abuse;[11] some types of mental disorders are potent risk factors for suicide (see below), and stress and anxiety are at the roots of the burnout syndrome seen in health workers (see below).

Extremely violent behaviour (and crime) is not associated with severe mental disorders (such as schizophrenia) as usually portrayed by the media.[12,13] On the contrary, it is more frequently associated with some types of personality disorders (e.g. antisocial, poor impulse control, aggressive) and with substance use, of which the most prevalent is alcohol. Hence, the control of substance use disorders (in itself a case of secondary prevention) can be seen as an example of primary prevention of domestic violence. The focus of prevention should be broadened from the individual perspective (the substance user) to the social group (the family) which is not exposed to the direct organic effects of alcohol, but nonetheless suffers from its indirect effects.

(b) Suicide

Despite the long-standing sociological tradition of considering suicide as a phenomenon associated with the social condition known as anomia,[14] the medical profession tends to see it primarily as a medical problem. Probably both are right. Among the demonstrated risk factors for suicide there are both social factors (anomy, old age, masculine gender, social isolation) and medical problems (chronic, painful, and incurable diseases and, most of all, psychiatric disorders and psychological problems). Several European studies have indicated that approximately 80 per cent of all cases of suicide are associated with alcohol use and depression combined.[2] This indicates the appropriateness of targeting the treatment of these two conditions (another secondary prevention intervention) regarding the prevention of suicide. However, given the broad social implications of suicide, treatment of mental disorders alone has not yet produced a significant reduction of suicide rates.

Hence the importance of a paradigmatic change which views suicide on a ecological perspective that considers the suicidal act as the immediate target for prevention and not just suicidal intention or ideation as subsumed by the traditional medical (and psychiatric) model.[2] It is now firmly established that the control of the availability of means to commit suicide can greatly contribute to the reduction of mortality rates from suicide. According to WHO,[2] steps to prevent suicide can be taken in the following areas:

- psychiatric treatment
- gun control
- gas detoxification
- control of toxic substances
- responsible media reporting.

Corresponding actions to prevent suicide include:

- identification and treatment of people suffering from depression
- restriction of access to guns
- detoxification of domestic gas and car emissions
- controlling the availability of toxic substances and medicines
- downtoning reports of suicide in the media
- erection of physical barriers to deter jumping from high places.

A more detailed discussion about the prevention of suicide can be found in Chapter 4.15.4.

(c) Staff burnout

Freudenberger[15] first used the term 'staff burnout' to describe a 'syndrome of exhaustion, disillusionment and withdrawal in voluntary mental health workers'. The concept has aroused considerable interest in the caring professions, and the publication of a large number of articles and books on the subject suggests that burnout is a major problem in health services.[2]

Indeed, with the general trend towards community-based care in many parts of the world, burnout is now a problem faced by all caregivers, including the relatives of people suffering from chronic disorders.

There is no single accepted definition of burnout; however, there is general agreement that the syndrome has three major characteristics, which are observed in various caregivers, particularly health workers and family members:

- emotional exhaustion
- depersonalization
- a reduced feeling of personal accomplishment.

The two approaches most frequently employed to prevent staff burnout are:

- Stress management (at the individual level);[16] and
- Supervisor training (at the organizational level).[17]

Table 7.4.2 summarizes actions which contribute to the prevention of burnout among health care staff, at different levels of intervention.

Who is responsible for primary prevention?

As indicated in the introduction of this section, one of the obstacles to a greater involvement of mental health professionals in preventive action stems from the fact that the effective actions for the

Table 7.4.2 Action to prevent staff burnout

Avoid making unrealistically high demands of caregivers
Ensure that all workers have some rewarding tasks
Train caregivers in time-management and relaxation techniques
Modify jobs that are proving too stressful
Encourage the formation of support groups
Consider the possibility of part-time employment
Encourage workers to participate in decisions which affect them

prevention of several mental disorders often falls outside their usual remit and in many cases falls outside the health sector altogether, as indicated in Table 7.4.3. Although this is true, it does not mean that mental health professionals do not have a major role to play in three areas: advocacy, information generation, and supervision.

Perhaps mental health professionals need to reconsider their potential role in primary prevention; for instance, they could develop their potential to act as advocates and advisers to professionals in other sectors. As Eisenberg[18] has argued:

> what matters is not the mode of action of the agent, the venue in which it is applied, or the academic discipline of the practitioner, but the effectiveness of the measure in preventing diseases manifested by disturbances in mental function.

At any rate, prevention is unquestionably a public health priority. Accordingly, WHO has recently published a book on this aspect of prevention, whose reading is strongly recommended.

Table 7.4.3 Agents for prevention

Condition	Effective agents for prevention
Mental retardation	
Down syndrome	Family planning professionals Obstetricians/midwives Women's associations
Foetal alcohol syndrome	General health-care personnel Substance abuse professionals Family planning professionals Obstetrictans/midwives . Educators Women's associations
Phenylketonuria	Obstetricians/midwives Paediatricians Nutritionists
Iodine deficiency	General health-care personnel Nutritionists Groups involved in salt production/trade Water supply personnel Educators
Depression	Mental health workers Rehabilitation officers
Schizophrenia	Mental health workers Rehabilitation officers
Alcohol dependence syndrome	General health workers Mental health workers Self-help groups
Violence	General health workers Mental health workers Paediatricians Justice officers Police officers
Suicide	Mental health workers General health workers Agricultural and environment authorities Journalists Pharmaceutical industry Car industry Traffic authorities Authorities in charge of provision of domestic gas Gun control authorities (including legislators)
Burnout	Occupational health workers General health workers Staff counsellors Trade unions Staff associations Personnel officers Job supervisors Self-help groups

Further information

Bertolote, J.M. (ed.) (1965). *Primary prevention of mental, neurological and psychosocial disorders*. World Health Organization, Geneva.

Leavell, H.R. and Clark, E.G. (1965). *Preventive medicine for the doctor in his community: an epidemiological approach* (3rd edn). McGraw-Hill, New York.

World Health Organization. (2004). *Prevention of mental disorders: effective interventions and policy options*. World Health Organization, Geneva.

References

1. Leavell, H.R. and Clark, E.G. (1965). *Preventive medicine for the doctor in his community: an epidemiological approach* (3rd edn). McGraw Hill, New York.
2. Bertolote, J.M. (ed.) (1998). *Primary prevention of mental, neurological and psychosocial disorders*. World Health Organization, Geneva.
3. Schneck, J.M. (1960). *A history of psychiatry*. C.C. Thomas, Springfield, IL.
4. Maj, M. (1990). Psychiatric aspects of HIV-1 infection and AIDS. *Psychological Medicine*, **20**, 547–63.
5. Gottesman, I. (1982). *The contribution of genetic factors to the common psychopathologies*. World Health Organization, Geneva.
6. Hetzel, B.S. (1986). Mental defect due to iodine deficiency: a major international public health problem that can be eradicated. In *Science and service in mental retardation* (ed. J.M. Berg), pp. 297–306. Methuen, London.
7. Dunn, J.T. and van der Haar, F. (1990). *A practical guide to the correction of iodine deficiency*. International Council for Control of Iodine Deficiency Disorders, Adelaide.
8. World Health Organization. (1985). *Community approaches to the control of hereditary diseases: report of a WHO advisory group*. World Health Organization, Geneva.
9. Smith, I.E., Lancaster, J.S., Moss-Wells, S., *et al.* (1987). Identifying high-risk pregnant drinkers: biological and behavioural correlates of continuous heavy drinking during pregnancy. *Journal of Studies on Alcohol*, **48**, 304–9.
10. Leff, J. and Waughn, R. (1981). The role of maintenance therapy and relative expressed emotion in relapse of schizophrenia: a 2-year follow up. *The British Journal of Psychiatry*, **139**, 102–4.
11. Regier, D.A., Farmer, M.E., Rae, D.S., *et al.* (1990). Comorbidity of mental disorders with alcohol and other drug abuse: results from the Epidemiological Catchment Area (ECA) study. *The Journal of the American Medical Association*, **264**, 2511–18.
12. Barbato, A. (1998). Psychiatry in transition: outcomes of mental health policy shift in Italy. *The Australian and New Zealand Journal of Psychiatry*, **32**, 673–9.

13. Aldige, H.V. (1992). Civil commitment and arrests: an investigation of the criminalization thesis. *The Journal of Nervous and Mental Disease*, **180**, 184–91.
14. Durkheim, E. (1990). *Le suicide*. Presse Universitaire de France, Paris.
15. Freudenberger, H.J. (1974). Staff burn out. *Journal of Social Sciences*, **30**, 159–65.
16. Meichenbaum, D. and Jaremko, M.E. (1983). *Stress reduction and prevention*. Plenum Press, New York.
17. Kilburg, R.C., Nathan, P.E., and Thoreson, R.W. (eds.) (1986). *Professionals in distress: issues, syndrome and solutions in psychology*. American Psychological Association, Washington, DC.

7.5

Planning and providing mental health services for a community

Tom Burns

Introduction

The aim of this chapter is to assist clinicians and managers review and plan services effectively for their local population. Severe psychiatric disorders manifest themselves in social relations and often disrupt social structures; they have wide-ranging consequences and services need to be comprehensive. Health and social care have been intertwined in psychiatry from its origins—it is neither feasible nor sensible to ignore the wider context of their management.

Mental health services research

The last 30 years have seen an explosion of Mental Health Services Research alongside the shrinking and closure of mental hospitals (see Chapter 7.6). Policy considerations, particularly cost containment and public safety, have influenced the research agenda which is disproportionately Anglophone (from the United States, United Kingdom, and Australasia) and focused on new services developed as alternatives to institutional care with staffing and motivation that are not easily generalizable. More routine practices, crucial for safe and effective care, have been relatively neglected by researchers.

Scope of chapter

This chapter is mainly devoted to describing the essential components of a mental health service—its 'building blocks'. It will then consider how they relate to one another, how they can be prioritized, and how integrated into an effective local service linking into other essential services. Lastly it will stress how their inevitable evolution should be monitored.

Services for adults (increasingly referred to as 'adults of working age' indicating 18–65 years) will be used as the template. In many settings these may be the only services, stretching to accommodate all comers. In better resourced health care systems a range of specialized services have evolved from this basic model and are described elsewhere in this section (refugees 7.10.1, homeless 7.10.2, and ethnic minorities 7.10.3).

Building blocks of mental health services: care and treatment

Most mental health treatments (whether psychological, pharmacological, or social) are based on face-to-face interviews and do not require sophisticated equipment or buildings. Institutions (the asylums) evolved for social care of disabled individuals, to protect them while they recovered and, sometimes, to protect society from them. Patients needing long-term institutional care are now relatively few but psychiatry is judged on how they are managed and service planners must pay them due attention.

Inpatient beds

No comprehensive service can survive without access to 24 h nursing supervision for acute episodes of severe illness. These include patients at risk from neglect or suicide or those lacking insight. Wards usually accommodate 10–20 patients. It is rarely possible to effectively staff and run stand-alone units of less than 3–4 such wards (30–60 beds). Ward size is a trade-off between privacy and domesticity against effective supervision. Single rooms are preferable, affording maximal privacy and, while initially expensive, improve flexibility and reduce conflict.

Smaller, more flexible, units such as 'crisis houses' offering 24 h care are a useful complement to inpatient wards, but not a replacement. Ward design and management are increasingly crucial as improved community care concentrates involuntary and disturbed inpatients in them.

How many acute beds?

'How many beds do we need for our local population?' is often the first question asked by planners or managers. Unfortunately there is no reliable or precise answer to this. We know that supply will drive use (perceived as need)—beds are rarely left unfilled despite enormous variation in their availability. It is also surprisingly difficult to collect useable figures on bed usage nationally or internationally because of differences in methods of reporting and also the profusion of overlapping and rarely defined local terms

(e.g. night hospitals, crisis homes, step-down wards). The levels of external accommodation provision (e.g. hostels, day care) clearly also impact the need for acute beds. Similarly need for beds will reduce as community services become more comprehensive and robust.

European provision of general acute beds in 2000 in the public sector ranged from 128 per 100 000 in the Netherlands to 6 per 100 000 in Northern Italy. However, unless we know the pattern of care (in particular the level of private and social services care) these figures tell us relatively little. The United Kingdom has little parallel private care and here acute beds needed for a population of 250 000 have been estimated to range from 50 to 150 plus 5 to 20 secure or intensive care beds[1] dependent on morbidity (generally much higher in large urban settings). London figures for the mid-1990s were very close to this range, averaging 73 for outer and 110 for inner London, but with increased secure provision, particularly in the deprived inner city. The authors predicted a similar range of 24 h supervised hostel need (40–150 per 250 000 population) and London use was somewhat higher (99 and 162 per 250 000, respectively) but with a markedly wider range.

Current bed usage in the United Kingdom is closer to the 50 per 250 000 and well below this in stable communities. This reflects both the establishment of specialized home-treatment and assertive outreach teams and the expansion of forensic care but also a shift in expectations and practice. The average duration for admissions has been steadily reducing over the last three decades. Figures can be misleading as they are heavily skewed by short (1–2 day) admissions but the current admission for an uncomplicated psychotic relapse is likely to be between 3 and 6 weeks.

Longer inpatient care

Acute inpatient wards admit patients for weeks or a couple of months. Rapid discharge is anticipated and regimes emphasize openness and independence. Even within a local service some patients will require longer or more secure care because of illness severity or for legal reasons. Modern rehabilitation practice restricts long-stay wards to patients whose behaviour is persistently unacceptable to local communities. Forensic and secure services are usually a regional or national rather than local responsibility.

Diagnosis-specific wards

Alcohol and substance abuse wards have been long established (especially in Scandinavia and Central Europe) and diagnosis- or disorder-specific wards are increasingly common. Wards for specialized patient groups such as anorexia nervosa or resistant schizophrenia provide highly specific regimes. These are generally an adjunct to acute admission wards rather than an alternative. Some services are organized in disorder-specific wards (e.g. a psychosis unit, a psychosomatic ward) *instead* of general wards. Such specialization is not possible in comprehensive services for populations of less than about 1 000 000. For smaller populations this increased specialization must be balanced against reduced flexibility and energy wasted in 'boundary disputes'.

Day care

Day care is provided either in day-hospitals or day-centres, with little consistency in the terms or practices. Patients attend usually from 1 to 5 days a week for a half or whole day before returning to their homes in the evening. It is particularly valuable when families are out at work but can offer support at evenings and weekends or for very isolated patients.

Generally day-hospitals are provided by health services, include medical and nursing staff and can offer treatments (e.g. the prescription and monitoring of medication, psychotherapies). Day-hospitals were a significant feature in the move of mental health services from mental hospitals to District General Hospital sites. However their role has been more uncertain since community teams have expanded and taken on much of their therapeutic role. Many services have scaled down or even closed their day-hospitals relying more on social services for day care. Day-hospitals have had a problem of isolating themselves from service needs, locked in time with a static patient group. Comprehensive services can, undoubtedly, survive well without them, so if they are to be established it is essential that there are strong links into local teams who can exercise some control over their clientele and their activity.

Day-centres, provided by social care organizations, can rarely provide treatments or employ clinical staff. However overlap is wide with services highly specific to local context (e.g. a drop-in day-centre may be the main provider of psychiatric assessment and treatment in areas of high social mobility and homelessness). Generally day-centres provide long-term social support and day-hospitals focused interventions and treatments.[2] The 'Club House' is a specialized rehabilitation day centre, popularized in the United States, which emphasizes useful normal work and where members take responsibility for running the centre with minimal supervision. Many day units now function in the evening and at weekends.

Acute day-hospitals in Europe and partial hospitalization in the United States have been energetically proposed as alternatives to inpatient care[3] but have had little impact. While day-hospitals never achieved their anticipated prominence they serve specific groups well (e.g. mothers with small children or protracted treatment of eating disorders or personality problems). Day care is problematic in rural settings but adaptations such as travelling day-centres (i.e. a team that moves from setting to setting on specific days) or a weekly open day run by the community team are worth considering.

Supported accommodation and residential care

Many patients remain well outside hospital only with adequate support. At its most basic this implies stable, affordable accommodation. For many, however, supervision is needed to ensure self care, continued medication, and to anticipate and defuse crises. This can be provided by voluntary agencies, social services, or health services. Voluntary agencies tend to be more efficient at providing long-term residential care[4] but they may be reluctant to accept risky patients (e.g. with a history of violence or substance abuse). A mixed economy works best and the need for health services supported accommodation depends on the vigour of local voluntary and social services. While some purpose built units exist, the accommodation is usually shared adapted houses to promote integration and reduce stigma.

Supported or sheltered accommodation is subject to a bewildering terminology but can be considered at four basic levels of increasing need:

1 ***Group homes***. These have no regular staff and are reserved for relatively independent patients visited by staff from their own community teams.

2 ***Day-staffed hostels***. One or two staff are present each day to support and monitor patients (encouraging cooking and cleaning, etc). They would usually not provide specific treatment but liaise with the community team about it.

3 ***Night-staffed hostels***. Non-clinical staff sleep over in the hostel to provide greater safety and availability.

4 ***24 h staffed/nursed hostels***. On-site clinical staff are available overnight either sleeping in or, sometimes, awake. These are expensive hostels and generally restricted to patients with long-term severe illnesses (including sometimes those compulsorily detained). Night-staffed hostels tend to be larger usually with 10–20 residents as opposed to 4–8 in day staffed ones.

Most comprehensive local services provide levels 1 and 2 and most social services undertake to provide level 3. Level 4 is relatively rare and would usually serve a population of 500 000–1 000 000.

Office-based care and outpatient clinics

In insurance-based systems many psychiatrists run individual office practices and manage patients on their own. In state-funded systems this is rare; most work in outpatient clinics or mental health teams. Both approaches should be considered when planning and providing public mental health services, paying particular attention to financial regulations that can inhibit integration and development (comprehensive planning may pose a significant threat to their livelihood and be resisted). Office-based practice remains widespread but neglected in academic and policy publications. It tends to be narrow in remit (usually either psychotherapy or pharmacotherapy) and is poorly equipped for managing severe disorders.

Outpatient clinics ('polyclinics' or 'dispensaries') are an essential part of modern services increasingly replacing office practice. Psychiatrists and psychologists may still operate independently within them but with access to enhanced resources and second opinions. In the public sector outpatient clinics may operate either alongside community mental health teams (CMHTs) or as part of them (which works better for severe illness).[5] They provide an efficient, predicable format for assessments, treatment, and monitoring.

Community mental health centres (CMHCs)

Mental hospitals, for all their faults, had no problems coordinating care; what little was available was all in the same place. Outpatient clinics expanded to Community Mental Health Centres (CMHCs) providing a wide range of services located in shared buildings (e.g. depot clinics, a day-hospital, psychotherapy services). The failure of the early US CMHCs demonstrated that relying entirely on patients to attend fails to engage the more ill and also that down-playing the 'medical model' made it impossible to recruit psychiatrists, further distancing practice from the severely ill.

Most CMHTs are based in CMHCs sharing accommodation with other CMHTs and services (e.g. day care). They provide an important safeguard in sustaining clinical standards and reducing the professional isolation in dispersed community services. This is a particularly important safeguard for community teams which can otherwise easily become idiosyncratic and rigid in their practice if not forced into regular contact with others.

Multidisciplinary Community mental health teams (CMHTs)

Most community mental health services consist of varied forms of multidisciplinary CMHT consisting of psychiatrists, nurses, social workers and often psychologists, and occupational therapists. The staffing of these teams will vary but their strength is that regular meetings to assess and review the management of patients incorporates their varied professional perspectives and allocates tasks based on skills and needs. Developed initially in France and the United Kingdom and championed latterly in Italy they have seen further specialization from North America and Australia.

The generic sector CMHT ('The CMHT')

Who it is for

The CMHT is *the* fundamental building block of modern community mental health services. It originated as mental hospital catchment areas (often covering a whole city or county) were divided into sectors of 50–100 000 inhabitants to permit ongoing care. The aim was that it should be possible for most of the team to have some familiarity with most of its complex and long-term patients and to have some *personal* knowledge of its referrers and community resources. Current sector size in Western Europe ranges from 20–50 000 population, determined both by resources (shrinking as investment increases) and by the local configuration. As more specialized teams are established the CMHTs remit may be reduced and sector size consequently increased keeping its caseload fairly constant. 200–250 is considered the maximum for most teams to exploit multidisciplinary working. The number is less in services for highly complex and difficult patients.

CMHTs offer assessment and care for patients discharged from psychiatric units and those who cannot be adequately treated in primary care or in the private sector. They should prioritize severe mental illnesses (SMI—e.g. psychoses and severe affective disorders). However diagnosis is not all—complications from social adversity, personality difficulties, or substance abuse can make secondary mental health care necessary even for apparently 'minor' disorders. Tools to clarify this threshold[6] have been of limited use and most teams rely on clinical assessments. In countries with limited private care CMHTs also treat mild and transient disorders. CMHTs can be remarkably inefficient if little thought is given to their structure and thresholds. To work well, there needs to be agreement on their purpose, clientele and systems of management and they have often suffered from lack of clarity and leadership.

Staffing and management

CMHT staffing varies enormously and there is no uniform model. Teams of less than 6 can rarely provide comprehensive care or cross-cover while teams of more than about 12–15 start to become

unwieldy, overwhelmed with management and information transfer. CMHTs emphasize skill-sharing and a degree of generic working and have evolved an informal, democratic style[7] which often means confusion over clinical leadership (originally provided informally by senior medical staff). With increased staff numbers and treatment complexity 'team managers' now coordinate workload with a role which varies from the purely administrative to setting clinical priorities and supervising staff. Establishing a clear understanding of clinical leadership in CMHTs (without inhibiting initiative and creativity) is essential for effective functioning. If leadership and management are separated (common with a strong medical presence) the roles need to be well defined and relationships good.

Assessments

The key to good care is accurate assessment (see Chapter 1.8.1). Most commonly psychiatrists conduct initial assessments (usually in an outpatient clinic) and involve the team members in treatment. Increasingly other team members have taken a role in assessments, either individually or jointly with the psychiatrist. Although this issue generates strong feelings there is surprisingly little research into it. With highly developed primary care non-medical assessments may be effective but otherwise medical time should prioritize assessments. With severely ill patients home-based assessments pay considerable dividends.[8,9]

Case management

Most CMHT staff act as clinical case managers[10,11] with responsibility for coordination, delivery, and review of care for their patients. The caseloads of staff members should be explicitly limited (usually 15 to 30) and reviews recorded and systematic. In the United Kingdom this has been formalized as the Care Programme Approach.[12] Fig. 7.5.1 shows a care plan indicating a patient's needs or problems, the interventions proposed to meet them, who is responsible and who is informed, plus an agreed date for review. Such concise structured paperwork (as with the risk assessment and contingency plan (Fig. 7.5.2)) can be adapted to any service, coordinates complex care and serves as a natural focus for clinical reviews. The level of detail needs to be clinically (not managerially) determined.

Team meetings

CMHTs need 1–2 regular meetings (each usually 1.5–2 h) per week for both clinical and administrative business. The degree of structure depends on team style and remit.

(a) Allocation of referrals

Referrals can be allocated by who is first available or by matching the clinical problem against available skill and training. Time discussing allocations before assessment is generally unprofitable and most well-established teams delegate the task to the manager or a senior clinician.

(b) Patient reviews

*R*eviews should be held for (i) *new patients*, (ii) *routine monitoring*, and (iii) *discharge*. Reviews can range from simply reporting the problem and proposed treatment in uncomplicated cases through to detailed, structured, multidisciplinary case-conferences including other services (e.g. GP, housing, child protection). *New patient* reviews are an excellent opportunity for providing a broad, experienced overview, and ensuring rational and fair allocation to caseloads. *Routine monitoring* is often overlooked yet probably the most important for team efficiency. It should be systematic and not only responsive to crises and problems. It shapes and redirects treatment and identifies patients ready for discharge. The burden on individual staff members is regularly monitored. Routine monitoring is a legal requirement of the Care Programme Approach and good practice in all case management. *Discharge reviews* are an excellent opportunity for audit and learning within the team.

(c) Managing waiting lists and caseloads

Effective CMHTs need to guarantee prompt access. *Routine assessments* should be within 2–4 weeks. Sooner is rarely productive and delays above 3 weeks result in a rapidly rising rate of failed appointments.[13] *Urgent assessments* (most psychotic episodes) need to be seen within a week, usually within a couple of days. *Emergency assessments* are for those associated with immediate risk (e.g. hostile behaviour or suicidal intent) and need to be seen the same day.

A practical approach to waiting lists is to count the assessments in the preceding year and allocate routine appointments for 20 per cent more. Thus a team with 400 assessments the preceding year allocating nine slots a week will have one available weekly for emergencies. Rapid routine assessment reduces pressure for urgent and emergency referrals more efficiently than emergency rotas.

Communication and liaison

Team meetings ensure internal communication but CMHTs need good links with a wide network of professional colleagues. Structured liaison is advisable with primary care and general hospitals in addition to routine letters. Hospital links may be between specific CMHTs and wards or CMHTs may provide input to patients from their sectors in the absence of dedicated liaison psychiatry services.

(a) General practice liaison

Much of mental health care is delivered in primary care (see Chapter 7.8) and effective coordination is essential. GP liaison systems range from informal contact through to shared care and co-location of CMHTs in GP Health Centres.[14] An effective system comprises regular, timetabled meetings between the two teams or a 'link' CMHT member attending the GP health centre. Monthly meetings where shared and complex patients are discussed are highly time-efficient because of prompt problem solving and crisis anticipation. However it is important to be clear about responsibilities, fudging boundaries is risky.

(b) Liaison with other agencies

The same principles apply to liaison with other agencies (social services, housing, charitable, and voluntary sector providers). Whether regular meetings are cost-effective will depend on the volume of shared work but showing up and meeting people (even just once) pays enormous dividends in improved relationships and understanding. Professional confidentiality and information sharing is more sensitive.

Assertive Outreach (AO) Teams

The most replicated and researched specialist CMHT is the AO Team. The original US model[9] improved clinical and social outcomes with substantially reduced hospitalization at slightly lower cost.

CPA REVIEW

Patient's name: Jenny T	**CMHT: West Central**
Address: 56 Acacia Avenue	**Phone:**
Phone:	**New patient: ~~YES~~/NO**
Date of birth: 09.06.61	**If NO, date of review: 20.10.07**
GP: Dr Findlay	**Diagnosis:**
Phone:	**1…Major depressive disorder..... F 32 .0**
	2…… F __ __.__

You must consider the following: 1) Mental health, including indicators of relapse; 2) Physical health; 3) Medication; 4) Daytime activity; 5) Personal care / living skills; 6) Carers, family, children and social network; 7) Forensic history; 8) Alcohol or substance misuse 9) Cultural factors; 10) Housing/finances/legal issues.

Complete a **risk assessment** and include: **i) a crisis plan; ii) a contingency plan**

Assessed needs or problem	Intervention	Resp.of
1. Depressed mood, apathetic and self critical	• Regular home visits, assess mental state • Encourage compliance with antidepressants • Encourage activity – take to shops etc	BJ
2. Suicidal thoughts	• Explore severity (+/- plans) at each visit • Support mother and husband who are scared of suicidal thoughts	BJ/ Cons
3. Daughter's school problems	• Maintain links with class teacher • Keep family informed of her progress	BJ
4. Plan for recovery	• Link with support group when mood lightens • Help reapply for part-time cleaning job	BJ

Professionals involved in care:	✓ Dr	Psychologist	✓ CPN	OT	✓ SW	Ward Nurse	ACT	Other
Present at planning meeting:	✓ Dr	✓ Psychologist	✓ CPN	OT	✓ SW	Ward Nurse	ACT	Other

Copy given to patient? **YES/~~NO~~** Copy sent to GP? **YES/~~NO~~**

Care co-ordinator(print): **Billie Jarvis (BJ)** **Phone**

Care co-ordinator (signature): ..... **Date of next review: 20.04.08.**

Job title: **CPN** **Patient's signature:……**

On Supervision Register? **~~YES~~/NO** Care management? **YES/~~NO~~** Risk history completed? **YES/~~NO~~**

On Supervised Discharge? **~~YES~~/NO** Relapse + risk plan required? **YES/~~NO~~**

Fig. 7.5.1 Care programme review document.

CONFIDENTIAL: RELAPSE AND RISK MANAGEMENT PLAN			
Name: Alastair W			
Categories of Risk Identified:			
Aggression and violence	~~YES~~/NO	Severe self-neglect	YES/~~NO~~
Exploitation (self or others)	YES/~~NO~~	Risk to children & young adults	YES/~~NO~~
Suicide and self-harm	YES/~~NO~~		
Other (please specify} ……			
Current factors which suggest there is significant apparent risk: **(For example: alcohol or substance misuse; specific threats; suicidal ideation; violent fantasies; anger; suspiciousness; persecutory beliefs; paranoid feelings or ideas about particular people)** Continued excessive drinking—especially when depressed. Makes him more suspicious and hostile.			
Clear statement of anticipated risk(s): **(Who is at risk; how immediate is that risk; how severe; how ongoing)** Clear risk to strangers (not family or staff), usually in bars. Often when poor medication compliance.			
Action Plan: **(Including names of people responsible for each action and steps to be taken if plan breaks down)** Relapse plan discussed and agreed—to increase antipsychotics and contact when concerned with people plotting ('to help you cope with them'). If he feels seriously threatened to seek admission through the emergency room			
Date Completed:	**xx/xx/xx**	**Review date:**	**xx/xx/xx**

Fig. 7.5.2 Risk assessment and contingency plan.

AO teams (Box 7.5.1) are costly and consequently reserved for the most difficult ('hard to engage' or 'revolving-door') psychotic patients with frequent, often dangerous, relapses and poor medication compliance plus alcohol or drug abuse, significant personality difficulties, and offending behaviour.

AO emphasizes proactive outreach—visiting patients at home even when they are reluctant. It exploits enhanced team working with daily meetings and several members actively involved with most patients both for safety considerations and also reflecting patients' extensive needs. The culture is of very practical working (taking patients shopping, sorting out accommodation, delivering medicines daily if need be) well beyond traditional professional boundaries.

Despite strongly expressed convictions there is little evidence that AO teams need to slavishly follow the original model[15,16] and local clinical adjustments are both sensible and justified. If embedded in a comprehensive system there is little need for a 24 h service, most staff establish strong individual relationships with patients and caseloads are usually more than the recommended 1:10. Where CMHTs function well AO teams take only patients who cannot be stabilized despite their support.

If CMHTs provide outreach and a comprehensive treatment then the extra AO resources may add little. Improved outcomes follow only from additional effective treatments (e.g. daily Clozapine visits in resistant schizophrenia); it is not outreach itself that is therapeutic. Whether AO will improve care (and, if so, how many teams are needed for how many patients) will depend on current services.

Box 7.5.1 ACT core components

- Assertive follow-up.
- Small caseloads (1:10–1:15).
- Regular (daily) team meetings.
- Frequent contact (weekly to daily).
- *In vivo practice* (treatment in home and neighbourhood).
- Emphasis on engagement and medication.
- Support for family and carers.
- Provision of services using all team members.
- Crisis stabilization 24 h a day, 7 days a week.

Ethics in community mental health care

Balancing patients' welfare with their autonomy and their rights with those of their families and the wider community are sharply revealed in AO teams. These teams regularly visit patients who vigorously and clearly reject them. When does intensive support become intrusion? When does professional persistence tip over into coercion or disrespect?

Compulsion was traditionally identified with the buildings of the old asylums or left to the family (as it still is in many parts of the world). With expanded community care, compulsion, and coercion (either explicitly in the form of legal requirements or informally through professional or social pressure[17]) are now a pervasive feature of practice. Improved legal and professional scrutiny makes compulsory treatment possible in the community. Most developed countries have enacted forms of community treatment order ('mandated community treatment', 'outpatient committal') mainly for the care of young psychotic individuals

without insight into their need for ongoing treatment. The introduction of these provisions has generally been controversial but their operation not so.

Community treatment orders have the advantage of legal scrutiny unlike most of the ethical dilemmas facing CMHT and AO staff in their day-to-day work. These require discussion case-by-case. How proper is it to inform neighbours if a patient may pose a risk to them but will not give consent? Is it right for a patient, heavily dependent on his parents, to deny them information on his treatment? What does a case manager do when they know their client is doing something illegal? Most professionals share the goal of maximizing their patient's autonomy while minimizing significant risks. Guidelines exist only for extreme circumstances. Teams should be encouraged in regular discussion of these issues.

Crisis teams

Crisis teams play a crucial role where local services are poorly developed (they may be the *only* community services) or in city centres with many transient and homeless patients. They must prioritize rapid response and accessibility. Most teams will see patients immediately, certainly the same day. Their clinical aims and staffing are essentially similar to the acute functioning in CMHTs. They are best located alongside CMHTs or in the emergency rooms of general hospitals with 24 h availability. Liaison services (see Chapter 5.7) can often incorporate and manage hospital-based crisis teams.

Crisis Resolution/Home Treatment (CR/HT) Teams

The CR/HT team model, developed in Australia[18] and currently implemented in the United Kingdom and Europe, reflects increased consumer demand for access in crises and a desire to reduce inpatient care costs. It draws heavily on AO practice with limited, shared caseloads, flexible working, extended access, and an emphasis on outreach. Reduction in hospitalization offsets much of their cost[19] but this needs to be considerable as with two daily shifts and on call overnight needs a staff of about 15 for a caseload of 30 proposed for a population of 150 000.[20] They target patients who would otherwise be in hospital and focus on the severely mentally ill with intensive visiting (usually daily for a limited period) and considerable practical support and work with patients' social networks; most aim for a maximum of 6 weeks involvement. Such intensive team working requires highly effective communication and the teams meet daily (often twice at shift handovers). Information transfer is burdensome and liaison with CMHTs complex requiring absolute clarity on local arrangements for clinical responsibility.

Variations in practice and sustainability

The CR/HT teams may reduce the need for hospital care[18,21] but how much will vary. The UK model is precisely specified (including who it should and should not care for, Box 7.5.2) but practice varies considerably. A full 24 h service is rarely needed, an on-call facility to the emergency room and police station at night usually suffices. Contact frequencies are generally lower, patients stay with the service longer than anticipated and they are inevitably referred individuals recurrently in crisis (often with alcohol and relationship problems) who cannot easily be refused care but would not be 'otherwise in hospital'. Good medical staffing is needed and CMHT responsibilities need to be carefully negotiated, mutually agreed and crystal clear if to be avoided. These realities need to be carefully considered before deciding to establish such teams.

Box 7.5.2 Remit of UK Crisis Resolution/Home Treatment Teams[20]

'Commonly adults (16 to 65 years old) with severe mental illness (schizophrenia, manic depressive disorders, severe depressive disorder) with an acute psychiatric crisis of such severity that, without the involvement of the CR/HT team hospitalisation would be necessary'.

'The service is not usually appropriate for individuals with:

- *Mild anxiety disorders.*
- *Primary diagnosis of alcohol or other substance abuse.*
- *Brain damage or other organic disorders including dementia.*
- *Learning disabilities.*
- *Exclusive diagnosis of personality disorder.*
- *Recent history of self harm but not suffering from a psychotic or serious depressive illness.*
- *Crisis related solely to relationship issues'.*

Crisis services may have a relatively limited lifespan[22] but can be a very successful way to improve local access, gain familiarity with at-risk populations, and then consolidate as a more comprehensive service. Sometimes, however, they become overwhelmed with inappropriate referrals or patients who cannot be referred on and close.

Crisis houses and respite care

Crisis houses allow admission with a minimum of formality and often with reduced supervision compared to hospitals. They are usually small (4–8 beds) in a domestic setting and take people for days, occasionally a week or two. They are favoured for vulnerable women and early intervention services. Most have one staff member sleeping in overnight and a couple on during the day with support from patients' case managers. They are very welcome for a minority of patients but do not replace inpatient care and need careful supervision to avoid becoming chaotic or blocked.

Adjunct or replacement for CMHTs?

The three teams outlined above comprise the fundamental building blocks of most community services. AO and CR/HT teams have been proposed as substitutes for CMHTs, particularly when there are problems with local CMHTs. However both experience and research evidence[23] suggests that they are rarely durable without effective CMHTs to relate to. They should be considered to improve the quality of care in otherwise well-functioning services rather than cost-saving shortcuts.

Highly specialized and diagnosis-specific teams

There are various specialized teams, generally organized on a regional level. These are not essential local services but impact on

them—both in terms of removing some of the clinical obligations and the need to ensure clear and negotiable thresholds.

Early Intervention Teams (EIS)

Concern that a long duration of untreated psychosis (DUP) confers poorer prognosis[24] has led to the development of EIS teams which many would now argue should be standard provision. Developed mainly from Australian and UK models[25,26] they vary remarkably even despite a detailed prescription for UK teams. Some down-play diagnosis in favour of easy access, others restrict to schizophrenia, some emphasize a 'youth service' while others take all first episode patients irrespective of age.[7] Even more confusing there are three quite different activities which may, or may not, be part of the service (Box 7.5.3).

The core of EIS is a specialized CMHT which case-manages first episode psychosis patients protecting social networks and functioning (keeping patients at college or work, an emphasis on family interventions, etc.) assuming a return to premorbid functioning. Crisis and respite houses are preferred to hospital. Some EIS teams conduct public awareness campaigns, lecturing in schools and colleges.[27] A minority of research teams attempt to identify and treat 'ultra-high risk' patients to prevent progression to psychosis.[28]

Forensic and rehabilitation teams

Community-focused services face particular difficulties in treatment-resistant patients, particularly those with socially unacceptable or offending behaviour. Such patients fit poorly into open wards and specialized forensic teams provide care where offending behaviour and danger to others predominates. Some provide community services (intensive case management of dangerous patients) with an emphasis on risk assessment and management. Integrating them with general services can be problematical.

Rehabilitation

A significant number of patients remain disabled despite best treatments and require long-term management of disability rather than episode-based care. Rehabilitation teams generally serve patients who cannot survive without supervised accommodation even when at their best. They include the diminishing cohort of old long-stay patients and increasingly a very disturbed 'new long-stay' population with comorbid substance abuse and behavioural disturbances.

Diagnostic-specific teams

Highly specialized teams for individual disorders (e.g. eating disorders, personality disorders, bipolar patients) concentrate specific skills and provide specialized treatments and are usually provided at regional level. They usually have stronger advocates than CMHTs, both from professionals and families of sufferers, and the opportunity costs (see below) of establishing them need careful thought.

Box 7.5.3 Components of early intervention teams

- **Case management**—ongoing care of identified patients
- **Early Identification**—awareness-raising campaigns for psychoses
- **High risk and prodromal** patient identification and treatment

Planning services

Step-wise planning and adaptation

Planning mental health services for a given community rarely starts with a clean sheet. In such circumstances there are excellent texts, both general and specific to mental health. Tansella and Thornicroft's 'matrix' model[29] is particularly thorough and structured (see Chapter 7.2). It covers the process from establishing service principles and needs assessment (at national, regional, and local levels) through to monitoring and reviewing the cycle of planning and provision. They propose a hierarchical approach depending on the level of mental health spend.[30] Case identification and outpatient treatments in primary care are the priority for low income countries and only with increased resources the establishment of a secondary care mental health service (usually a form of generic CMHT). Not until these are well-established are specialist and inpatient services indicated. This process must, however, take account of what is already in place.

Local population needs assessment

Psychiatric morbidity varies considerably with social deprivation and is much higher in cities than in stable rural or suburban settings (see Chapter 2.7). At the regional and national level comparative need can be predicted fairly well from established indexes incorporating levels of migrants, overcrowding, poverty, etc. Catchment areas should broadly reflect these differences. At the more local level these figures are of limited value. How does one factor in travelling time or known differences in the quality of primary care? A process of negotiation is best to agree local allocation following these general guidelines. However a concentration of hostels for the mentally ill or homeless or the presence of a railway station or international airport may swamp these differences. These should be provisionally estimated in planning but then regularly monitored and reviewed.

Opportunity costs and unintended consequences

Planning mental health services has become based on international evidence, often including cost-effectiveness analyses, i.e. is there more overall patient health gain for the same input (see Chapter 7.7)? However these rarely address the opportunity costs across a whole system. For example one form of day hospital may be more cost-effective for its patients than another day hospital, but is diverting nurses from an inpatient unit to staff that day hospital a net gain? Rigorous intervention studies of large systems are formidably difficult to conduct and even more so to interpret.

The impact of enthusiasm and the migration of the best staff to such research services can be especially misleading.[31] Successful new services are always reported but there is less consistency in reporting when a service may have lost its efficacy or was abandoned.[5,22] It is best to visit examples of services that are proposed and not to always assume that what works for them will necessarily work for you.[15]

Manpower is often as significant a resource limitation as funding, hence the need for a system-wide appraisal. Also, though expressed as costs, these decisions include wider judgements of values and expectations rather than simply outcomes.[30] Local and national objectives also matter, not just clinical ones (e.g. public safety now dominates much mental health planning).

Cultures and funding

Health care cultures, their structures and their funding systems vary enormously. Services must be congruent with them otherwise they will not survive. Occasionally service planners can influence the system but more often have to adapt to it. Obtaining relatively small changes in funding arrangements (or external governance) can deliver quite major improvements. However caution is needed as unintended consequences and perverse incentives may arise. Enthusiastic clinicians are often blind to the risks of over-prescription which can lock in outmoded practices (e.g. a highly specific form of day hospital or crisis facility) that after opening is found not to have the level of need predicted but may be so rigidly prescribed that it cannot be adapted.

How 'integrated' mental health care should be is a highly local decision. Well-established systems often strive for integration with general medicine or, increasingly, with social services to reduce both discrimination and administrative barriers to integrated care. For less confident services the value of a distinct, separate, identity can be considerable—not least the ability to protect its resources. The 19th century British mental health reformer Lord Shaftsbury wrestled with the same dilemma.

The task of integration is easily underestimated, particularly the energies required to accommodate contrasting health and social care cultures. Social care, for example, tends to rely on very detailed paperwork and a highly structured system of intensive supervision (reflecting political accountability and previously only minimally trained workforce) as opposed to the high levels of professional autonomy in health care. Agreed compromises need to be reached before integration and the negotiations can be exhausting, but preferable to misunderstandings. The costs and benefits will depend very much on the local situation, relationships and history and should be very carefully weighed up.

Relationships with the voluntary sector and patients movement

Relations with local non-health statutory services (housing, education, police) and with the voluntary sector will determine much of the success of MH services. Voluntary and private sectors may fill specialist niches left by a monolithic public health care system (as with the NHS in the United Kingdom) or conversely the public system may act as a safety net, for charitable and private provision. Service planners need to exploit the strengths of local providers. Non-statutory services are often more efficient but less comprehensive and may also be less reliable over the long-term. Both may be equally reliant on public funding, differing only in contracts and degrees of independence. Patient and carer advocacy and support organizations are now a major force; effective working with them significantly will enhance both the design and delivery of services (see Chapter 7.9).

Monitoring and review

Careful monitoring and review are as important as careful planning for several reasons. The long-term, fluctuating nature of disorders, the subjective nature of diagnoses (increasingly self-ascribed) complicate outcome measures and the targeting of services; services can easily drift to those who demand them from those who most need them; treatments (particularly psychological and psychotherapeutic treatments) may evolve over time losing their effective characteristics; needy patients are rarely demanding or well-informed; engaging and motivating them to collaborate in long-term, treatments is difficult.

A consistent effort to deliver even the most basic, proven interventions will make a substantial difference to patient welfare. The PORT study in the United States demonstrated how schizophrenia care was strikingly inconsistent and fell below accepted essentials.[32] Regular audit ensures that services remain targeted on those for whom they were developed and that their application and quality remains good. Monitoring can vary in sophistication—from a simple head count of who is getting what through to careful evaluation of care pathways. The audit review process feeds back into the development process, adjusting and refining it. Even in the most hard-pressed services audit more than rewards its investment.

Routine outcome measures (ROM)

Audit brings rigour and reflection (often lost in the immediacy of the therapeutic relationship) into the care process. Measurement also serves a training purpose by benchmarking interdisciplinary understandings of symptoms and outcomes. Systematic, periodic recording of patients' clinical or social status is increasingly used in both planning and research. Structured outcomes can be generic (e.g. HoNOS[33]) or for specific disorders (e.g. the Brief Psychiatric Rating Scale[34]) or even locally developed. Their value lies in their consistency of use.

Conclusions

Planning and providing mental health services requires flexibility and compromise. Epidemiological and service statistics translate poorly to local planning; this remains primarily a practical and political, rather than academic, activity. The scientific evidence is dominated by Anglophone alternatives to hospital care studies but local history, culture, mental health law, and political imperatives cannot be ignored. Mediterranean societies, for example, with strong family supports and fewer isolated psychotic individuals have less interest in AO teams. A high-profile patient homicide, or a strong public endorsement of services by politician or celebrity, can derail years of careful planning.

This chapter has attempted to draw out some principles (Box 7.5.4) for the process but these can only be guidelines. Inevitably the decision will be based on what is *possible* locally. Despite this most solutions draw on a limited number of tried and tested structures described here in some detail. Their balance and configuration depend on what is available for them and around them. Above all the mentally ill and their families deserve reliable and predictable services. Not all change is innovation and research findings should be judged in terms of their sustainability and demonstrated translation from research efficacy to clinical effectiveness.

Box 7.5.4 Developing local community mental health services

- Make a careful inventory of what services exist, and any special local needs.
- Consult locally and invest heavily in building coalitions with policy makers, statutory services, and voluntary groups.
- Test research evidence for durability and relevance. If possible visit established services and ask 'around the service'.
- Monitor and review regularly. Improved consistency of current practice often delivers more than introducing new treatments.
- Consider carefully opportunity costs. Include both the impact of a specific improvement across the whole service and the costs of system change itself.
- Avoid excessive reorganization—not all change is innovation.

Further information

Burns, T. (2004). *Community mental health teams*. Oxford University Press, Oxford.

The Mental Health Policy Implementation Guide. (2001). Department of Health, London.

Thornicroft, G. and Tansella, M. (2004). Components of a modern mental health service: a pragmatic balance of community and hospital care: overview of systematic evidence. *The British Journal of Psychiatry*, **185**, 283–90.

Thornicroft, G. and Tansella, M. (1999). *The mental health matrix: a manual to improve services*. Cambridge University Press, Cambridge.

References

1. Strathdee, G. and Thornicroft, G. (1992). Community sectors of need-lead mental health services. In *Measuring mental health needs* (eds. G. Thornicroft, C.R. Brewin, and J. Wing). Gaskell, London.
2. Catty, J., Goddard, K., and Burns, T. (2005). Social services day care and health services day care in mental health: do they differ? *International Journal of Psychoanalysis*, **51**(2), 151–61.
3. Marshall, M. (2003). Acute psychiatric day hospitals. *British Medical Journal*, **327**, 116–7.
4. Knapp, M., Hallam, A., Beecham, J., *et al.* (1999). Private, voluntary or public? Comparative cost-effectiveness in community mental health care. *Policy and Politics*, **27**(1), 25–41.
5. Wright, C., Catty, J., Watt, H., *et al.* (2004). A systematic review of home treatment services. Classification and sustainability. *Social Psychiatry and Psychiatric Epidemiology*, **39**, 789–96.
6. Slade, M., Powell, R., Rosen, A., *et al.* (2000). Threshold assessment grid (TAG): the development of a valid and brief scale to assess the severity of mental illness. *Social Psychiatry and Psychiatric Epidemiology*, **35**(2), 78–85.
7. Burns, T. (2004). *Community mental health teams*. Oxford University Press, Oxford.
8. Burns, T., Beadsmoore, A., Bhat, A.V., *et al.* (1993). A controlled trial of home-based acute psychiatric services. I: clinical and social outcome. *The British Journal of Psychiatry*, **163**, 49–54.
9. Stein, L.I. and Test, M.A. (1980). Alternative to mental hospital treatment. I: conceptual model, treatment program, and clinical evaluation. *Archives of General Psychiatry*, **37**(4), 392–7.
10. Intagliata, J. (1982). Improving the quality of community care for the chronically mentally disabled: the role of case management. *Schizophrenia Bulletin*, **8**(4), 655–74.
11. Holloway, F., Oliver, N., Collins, E., *et al.*(1995). Case management: a critical review of the outcome literature. *European Psychiatry*, **10**, 113–28.
12. Department of Health. (1990). The care programme approach for people with a mental illness referred to the special psychiatric services. Department of Health, London. Report No.: Joint Health/Social Services Circular HC (90) 23/LASS (90) 11.
13. Burns, T., Raftery, J., Beadsmoore, A., *et al.* (1993). A controlled trial of home-based acute psychiatric services. II: treatment patterns and costs. *The British Journal of Psychiatry*, **163**, 55–61.
14. Burns, T. and Bale, R. (1997). Establishing a mental health liaison attachment with primary care. *Advances in Psychiatric Treatment*, **3**, 219–24.
15. Fiander, M., Burns, T., McHugo, G.J., *et al.* (2003). Assertive community treatment across the Atlantic: comparison of model fidelity in the UK and USA. *The British Journal of Psychiatry*, **182**, 248–54.
16. Burns, T., Marshall, M., Catty, J., *et al.* (2005). *Variable outcomes in case management trials—an exploration of current theories using meta-regression and meta-analysis*: Final Report. Department of Health, London.
17. Monahan, J., Redlich, A.D., Swanson, J., *et al.* (2005). Use of leverage to improve adherence to psychiatric treatment in the community. *Psychiatric Services*, **56**(1), 37–44.
18. Hoult, J. (1986). Community care of the acutely mentally ill. *The British Journal of Psychiatry*, **149**, 137–44.
19. Smyth, M.G. and Hoult, J. (2000). The home treatment enigma. *British Medical Journal*, **320**(7230), 305–9.
20. Department of Health. (2001). The mental health policy implementation guide. Department of Health, London.
21. Johnson, S., Nolan, F., Pilling, S., *et al.* (2005). Randomised controlled trial of acute mental health care by a crisis resolution team: the north Islington crisis study. *British Medical Journal*, **331**(7517), 599.
22. Cooper, J.E. (1979). Crisis admission units and emergency psychiatric services. Public Health in Europe, No. 2. Copenhagen: World Health Organisation.
23. Burns, T., Catty, J., Watt, H., *et al.* (2002). International differences in home treatment for mental health problems. Results of a systematic review. *The British Journal of Psychiatry*, **181**, 375–82.
24. Marshall, M., Lewis, S., Lockwood, A., *et al.* (2005). Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review. *Archives of General Psychiatry*, **62**(9), 975–83.
25. Edwards, J., McGorry, P.D., and Pennell, K. (2000). Models of early intervention in psychosis: an analysis of service approaches. In *Early intervention in psychosis: a guide to concepts, evidence and interventions* (eds. M. Birchwood, D. Fowler, and C. Jackson). John Wiley & Sons, New York.
26. Birchwood, M., Todd, P., and Jackson, C. (1998). Early intervention in psychosis. The critical period hypothesis. *The British Journal of Psychiatry—Supplement*, **172**(33), 53–9.
27. McGorry, P. and Jackson, H. (1999). Recognition and management of early psychosis. A preventative approach. Cambridge University Press, Cambridge.
28. McGorry, P.D., Yung, A.R., Phillips, L.J., *et al.* (2002). Randomized controlled trial of interventions designed to reduce the risk of progression to first-episode psychosis in a clinical sample with subthreshold symptoms. *Archives of General Psychiatry*, **59**(10), 921–8.
29. Tansella, M. and Thornicroft, G. (1998). A conceptual framework for mental health services: the matrix model. *Psychological Medicine*, **28**(3), 503–8.

30. Thornicroft, G. and Tansella, M. (2004). Components of a modern mental health service: a pragmatic balance of community and hospital care: overview of systematic evidence. *The British Journal of Psychiatry*, **185**, 283–90.
31. Coid, J. (1994). Failure in community care: psychiatry's dilemma. *British Medical Journal*, **308**(6932), 805–6.
32. Lehman, A.F. and Steinwachs, D.M. (1998). Translating research into practice: the Schizophrenia Patient Outcomes Research Team (PORT) treatment recommendations. *Schizophrenia Bulletin*, **24**(1), 1–10.
33. Orrell, M., Yard, P., Handysides, J., *et al.* (1999). Validity and reliability of the health of the nation outcome scales in psychiatric patients in the community. *The British Journal of Psychiatry*, **174**, 409–12.
34. Overall, J.E. and Gorham, D.L. (1962). The brief psychiatric rating scale. *Psychological Reports*, **10**, 799–812.

7.6

Evaluation of mental health services

Michele Tansella and Graham Thornicroft

Introduction

Evaluation is the basis for improving care to people with mental illness. It is vital to know whether interventions are beneficial or harmful, and whether they offer value for money. Mental health interventions need to be understood both in terms of their active ingredients and how they fit within their context.[1] Such combined interventions, often including pharmacological, psychological, and social elements, are the epitome of 'complex interventions'[2] and their evaluation poses considerable challenges. In this chapter we shall discuss *definitions* of evaluation, and go on to discuss *why* evaluate, *what* to evaluate, and *how* to evaluate mental health services. In our conclusion we shall offer an indication of the most important trends in this field in the coming years. The overall approach that we take is centred upon the idea that ongoing evaluative research is of fundamental importance in discovering which interventions are effective, neutral, or harmful, and that such information is essential to deliver better mental health care.

Evaluation: definitions and conceptual framework

The *Concise Oxford English Dictionary*[3] gives the following definitions of 'evaluation':

evaluate (*verb transitive*) **1**. assess, appraise; **2a**. find or state the number or amount of; **2b**. find a numerical expression for.

evaluation (*noun*) **1**. appraisal, valuation, assessment; **2**. estimate, estimation, approximation, rating, opinion, ranking, judgement, reckoning, figuring, calculation, computation, determination.

The etymological root of the word therefore refers directly to 'value', although in common usage 'evaluation' now has a more technical connotation. In our view evaluation necessarily requires both the precise measurement of the effects of treatments or services, alongside a contextual understand of the meaning, and value of such results.

A conceptual model that can be used to clarify key issues related to the evaluation of mental health services is the Matrix Model.[4,5] The two *dimensions* of this model are place and time (see Table 7.6.1). Place refers to three geographical levels: (1) country/regional, (2) local, and (3) individual. Time refers to three phases: (A) inputs, (B) processes, and (C) outcomes. In this framework *inputs* relates to all those resources which are necessary before health care can take place (such as financial and human resources, policies, and treatment guidelines), *processes* refers to all those activities which constitute the delivery of health care (such as outpatient consultations, or hospital admissions), while *outcomes* refers to the consequences of health care (such as changes in symptoms, disability, and quality of life). In relation to the evaluation of mental health services, we shall illustrate in this chapter how inputs and processes need to be measured and understood in their contribution to the outcomes of care.

Historically, the first attempts to evaluate psychiatric practice originated in the mid-nineteenth century as the tabulation of admissions, discharges, and deaths in mental hospitals, simply describing the inputs and processes of care. In recent decades, as more sophisticated research methodologies and more valid and reliable research measures have been developed, so the evaluation of mental health services has increasingly focussed upon the analysis of the outcomes of care. As Sartorius has put it, 'In its most classical form, evaluation denotes a comparison between results and goals of activity',[6] indicating that evaluation has now become a purposeful exercise in which measurements are used as tools to answer specific questions, usually defined a priori at the beginning of a scientific study.

Why evaluate mental health services?

In our view, the main purposes of mental health service evaluation are to assess the effectiveness and cost-effectiveness of care, either at the organizational (local) or at the patient (individual) level. In the long-term such evidence can be used to provide better services for people with mental illness. For example, evaluation can be applied to comparing differing models of care, such as studies in England showing that home-treatment teams can provide a realistic alternative to emergency hospital admission.[7–9] Evaluation therefore measures the impact of care (outcomes) and also aims to increase understanding of the active ingredients (inputs and processes) which contribute to better outcomes.[1] In fact, a wider range of purposes can be served by the evaluation of mental health services, as shown in Table 7.6.2.

What to evaluate in mental health services?

In our view the most important focus of evaluation is upon the *outcomes* of care.[10,11] The outcome chosen for any particular

Table 7.6.1 Overview of the Matrix Model, with examples of inputs, processes, and outcomes

Place dimension	Time dimension		
	(A) Input phase	(B) Process phase	(C) Outcome phase
(1) Country/regional level	1A Mental health budget allocation Mental health laws Government directives and policies Training plans for mental health staff Treatment protocols and guidelines	1B Performance/activity indicators (e.g. admission rates, compulsory treatment rates)	1C Overall suicide rates Homelessness rates Imprisonment rates Years lived with disability
(2) Local level	2A Local service budgets and balance for hospital and community services Local population needs assessment Staff numbers and mix Clinical and non-clinical services Working relationships between teams	2B Service contacts and patterns of service use Pathways to care and continuity Targeting of services to special groups	2C Suicide rates among people with mental illness Employment rates Physical morbidity rates
(3) Individual level	3A Assessments of individual needs made by staff, service users, and by families Therapeutic expertise of staff Information for service users Information for family members	3B Content of therapeutic interventions (both psychological, social, and pharmacological) Continuity of clinical staff Frequency of appointments	3C Symptom severity Impact on caregivers Satisfaction with services Quality of life Disability Met and unmet needs

evaluation will depend upon the central question addressed and the level at which outcomes are assessed, as shown in Table 7.6.3.

Directly in relation to the population level, a frequently used outcome measure is suicide rate (see cell 1C in Table 7.6.1). Rates of homelessness among mentally ill people (or rates of mental illness among the homeless) can also be used as an outcome indicator of the effectiveness of mental illness policies at the national (or regional) level.

At the local level, outcome indicators useful for evaluation can be made in three ways: (i) by interpolating from regional/national data; (ii) by measuring directly at the local level; and (iii) by aggregating individual-level information up to the local level. For example, rates of suicide and unemployment can be estimated using the first method, or directly measured using the second approach if the appropriate data and resources exist, which will provide more accurate and up-to-date information. The third approach is to aggregate up to the local level information gathered from individual patients, if institutions providing care to those local patients are willing to cooperate in integrating their datasets.

At the individual level mental health service evaluation increasingly acknowledges the importance of outcomes other than symptom severity.[10,11] Traditionally, **symptom severity measures** have been used most often to assess the effectiveness of the early, mental health treatments. Psychiatrists and psychologists have contributed to the early development of such assessment scales to allow this

Table 7.6.2 Main purposes of mental health service evaluation

- To assess the outcomes of services in experimental conditions (*efficacy*)
- To investigate whether interventions which have demonstrated *efficacy* under experimental conditions are also *effective* in ordinary, routine clinical conditions
- To understand the mechanism of action (i.e. active ingredients) of interventions
- To inform mental health service investment decisions, for example using health economic data on cost-effectiveness
- To raise awareness among planners, policy makers, and politicians of service gaps
- To test a priori or to check *post hoc* the value of planning decisions (for example, the closure of mental hospitals)

Table 7.6.3 Outcome measures suitable for use in routine clinical practice

Outcome measure	Place dimension		
	Country level	Local level	Individual level
Employment status	√	√	√√
Physical morbidity	√	√	√
Suicide and self-harm	√√	√	√√
Homelessness		√	√√
Standardized mortality ratios	√	√	
Symptom severity		√	√√
Impact on caregivers		√	√
Satisfaction with services		√	√√
Quality of life		√	√
Disability		√	√√
Met and unmet needs for care		√	√

Key: √ = suitable for use as an outcome, √√ = commonly used as an outcome.

research to take place.[10,11] While the primary symptoms are clearly important, for most of the more severe mental disorders there is symptom persistence, and, at present, it is unrealistic to see symptom eradication as the sole aim of treatment. Therefore, very often, after the point of maximum symptom relief, when the extent of the ongoing impairments is clear, then the clinical task becomes one of attempting to minimize the consequent disability and handicap.

The importance of the **impact of caring** for people with mental illnesses upon family members and others who provide informal care has long been recognized, but has only been subjected to concerted research relatively recently.[11–13] Such research has shown that it is common for carers themselves to suffer from mental illnesses, most commonly depression and anxiety, and to worry about the future when they may no longer be able to cope. Moreover, many family members are most distressed by the patient's underactivity, and are often poorly informed about the clinical condition, its treatment, and the likely prognosis, as well as being inadequately provided with a practical action plan of what to do in the future should a crisis occur. Indeed, some services continue to convey to families the outmoded idea that carers, especially parents, are in some way to blame for the disorder or for relapses of the condition. The regular provision of information sessions for family members is now a hallmark of a good practice.[14,15]

Patients' **satisfaction with services** is a further domain that has recently become established as a legitimate, important, and feasible area of outcome assessment.[16] This is a recognition of the contribution that service users and their carers can make to outcome assessment. Psychometrically adequate scales in this field are those that adopt a multidimensional approach, assess the full range of service characteristics, are independently administered (so that patient ratings have no consequences upon their future clinical care), and have established validity and reliability.[17]

Quality of life ratings have also become prominent during the last decade, and several scales have been constructed that reflect different basic approaches to the topic.[18] The first distinction is between scales that address subjective well-being, compared with those that also measure objective elements of quality of life. The second main point of differentiation is between scales constructed for the general population and those designed for patients suffering from specific disorders, including the more severe mental illnesses.[19] One advantage of quality of life data is that they tend to be popular with politicians, for whom the concept often has powerful face validity.

Among people with longer-term or more complex mental illnesses, the measurement of **disability** is often an important consideration.[20] Increasing importance is also being attached to the **needs** of people with mental illness, where met needs are difficulties faced by people with mental illness in the presence of appropriate interventions.[21] Needs (both met and unmet) may be defined by professionals/experts, or by service users, and in fact there is emerging evidence that service user ratings may be more informative, for example in predicting quality of life.[22–24]

Psychometric properties of outcome measures

Establishing the psychometric qualities of scales used for service evaluation is a central issue.[4] Among the most important characteristics of outcome scales are validity and reliability. Validity refers to whether a scale actually measures what it is intended to measure. It is conventionally assessed in terms of face validity, content validity, consensual validity, criterion-related validity, and construct validity.

In addition, a rating scale must give repeatable results for the same subject when used under different conditions, i.e. it must be reliable. There are four widely used methods to gauge reliability: inter-rater reliability, test–retest reliability, parallel-form reliability, and split-half reliability. The main issue for the evaluation of mental health services is to use wherever possible scales with known and adequate psychometric properties.

How to evaluate mental health services

In this section we consider research designs that may be applicable to the range of contexts used in mental health service evaluation.[1] Different types of evidence produced using these designs cannot be considered as equivalent. A hierarchical order has been proposed by Geddes and Harrison[25] as shown in Table 7.6.4.

In terms of research methods or designs which can be used to produce such evidence, they can be considered as: (i) randomized controlled trial (RCT), (ii) quasi-experimental studies, (iii) case-control studies, (iv) cohort studies (prospective or retrospective), (v) cross sectional studies, and (vi) case series and single case studies. Since evaluations of mental health services are usually concerned with complex interventions, it is helpful to have an overall scheme linking different stages of research to test treatment interventions. The Medical Research Council (MRC) framework for the evaluation of complex interventions sets out one such sequence, as shown in relation to anti-stigma interventions in Table 7.6.5. The elements in this scheme can be considered as sequential, or stages 0, 1, and 2 can be seen as one larger iterative activity.[1] Nevertheless, although this gives salience to randomized controlled trial designs, it is important to appreciate that research study designs need to be matched to the purpose of each type of evaluation, as shown in Table 7.6.6.

Evidence from a meta-analysis of randomized controlled trials

Meta-analysis can be defined as 'the quantitative synthesis of the results of systematic overviews of previous studies', while systematic overviews, in turn, are methods of collating and synthesizing all the available evidence on a particular scientific question.[26] Since randomized controlled trials are often considered to produce the most sophisticated evidence on the efficacy of medical treatments,

Table 7.6.4 Hierarchy of evidence

1a	Evidence from a meta-analysis of RCTs
1b	Evidence from at least one RCT
2a	Evidence from at least one controlled study without randomization
2b	Evidence from at least one other type of quasi-experimental study
3	Evidence from non-experimental descriptive studies, such as comparative studies, correlation studies, and case-control studies
4	Evidence from expert committee reports or opinions and/or clinical experience of respected authorities

(Reproduced from J.R. Geddes, and P.J. Harrison, Closing the gap between research and practice, *The British Journal of Psychiatry*, **171**, 220–5, copyright 1997, The Royal College of Psychiatrists.)

Table 7.6.5 Phases of the Medical Research Council framework for the evaluation of complex interventions[1,2]

0 Preclinical	1 Modelling/manualization	2 Exploratory	3 Definitive trial	4 Long-term implementation
Explore relevant theory to ensure best choice of intervention and hypothesis and to predict major confounders and strategic design issues	Identify the components of the intervention and the underlying mechanisms by which they will influence outcomes to provide evidence that you can predict how they relate to and interact with each other	Describe the constant and variable components of a replicable intervention and a feasible protocol for comparing the intervention with an appropriate alternative	Compare a fully defined intervention with an appropriate alternative using a protocol that is theoretically defensible, reproducible, and adequately controlled in a study with appropriate statistical power	Determine whether others can reliably replicate your intervention and the results in uncontrolled settings over the long-term
Example: anti-stigma intervention in schools study				
Social contact theory[69]	Yes	Completed[70]	Planned	Potential if preceding phases successful

a meta-analysis conducted on well selected and relevant randomized controlled trials can be seen as the highest order of knowledge. It follows that the quality of systematic overviews is limited by the quality and quantity of the contributory trials (see Table 7.6.7).[27]

Cochrane was the first to emphasize the need to bring together, within specific categories, the results of randomized controlled trials.[28] This approach is now central to evidence-based medicine. Within psychiatric evaluation the first meta-analyses were conducted in the late 1970s, and more information is given on systematic reviews in Chapter 1.10 and 6.1.

An illustration of such an exercise is the systematic overview and meta-analysis is that which reviewed RCTs comparing the outcomes of community mental health teams with those of standard care for patients with severe mental illness and disordered personalities.[29,30] They found 1200 citations using the search strategy: 70 appeared relevant to the review, but only four studies satisfied the inclusion criteria. The main results of this systematic review are that community mental health team management is associated with fewer deaths by suicide, with fewer people being dissatisfied with services or leaving the studies early. No clear difference was found in admission rates, overall clinical outcomes, or in the duration of inpatient hospital treatment. The authors concluded that community mental health team management is not inferior to non-team standard care in any important respects, and is superior in promoting greater acceptance of treatment. It may also be superior in reducing hospital admissions and avoiding deaths by suicide.

The randomized controlled trials

The importance of randomized controlled trials within medical research has been expressed by Korn and Baumrind:[31] 'Randomized clinical trials are the sine qua non for evaluating treatment in man'. According to Barker and Rose: 'the essence of the randomized controlled trial is that the outcome of the treatment given to one group of patients is compared with one or more other groups who are given different treatments or none at all. Allocation of individuals to the treatment and comparison groups is by random selection'.[32]

Table 7.6.6 Research aims and appropriate study designs

Research aim	Appropriate study designs
Service description	Cross sectional survey
Assess intervention	Quasi-experimental study (e.g. controlled before–after comparison) Randomized controlled trial
Identify prognosis for a condition	Cohort study
Establish aetiology of a condition	Cohort study Case-control study

The advantages of the research design of these studies have been extensively described[33] and are shown in Table 7.6.8 (see also Chapter 1.10 in this book). However, the design limitations of such trials also need to be appreciated, particularly in relation to health service research, as shown in Table 7.6.9.[34]

In addition to the technical limitations of the trial design, there are also situations where randomized controlled trials are not applicable to specific research questions. These can be summarized as conditions in which randomized controlled trials are inadequate, impossible, inappropriate, or unnecessary, as shown in Table 7.6.10. Nevertheless, where they are appropriate, it will be necessary to use explicit criteria to assess the quality of such trials, such as those shown in Table 7.6.11.[35,36]

It is now common to distinguish between *efficacy trials* (which tend to be explanatory) and *effectiveness trials* (sometimes otherwise called large simple, pragmatic, practical, or management trials).[33,37–39] This categorical distinction has its uses, although

Table 7.6.7 Characteristics of systematic overviews

Questions to ask about papers for potential inclusion in a systematic overview

- Were the questions and methods clearly stated?
- Were comprehensive search methods used to locate the relevant articles?
- Were explicit methods used to determine which articles were included in the review?
- Was the methodological quality of the primary studies assessed?
- Were the selection and assessment of the primary studies reproducible and free from bias?
- Were the differences in individual study results adequately explained?
- Were the results of the primary studies combined appropriately?
- Were the reviewer's conclusions supported by the data cited?

(Reproduced from S.I. Sackett and J.E. Wennberg, Choosing the best research design for each question, *British Medical Journal*, **315** (7123), 1636, copyright 1997, BMJ Publishing Group Ltd.)

Table 7.6.8 Advantages of RCTs

- Controls for many confounding variables which may exist
- Eliminates the effects of spontaneous remission
- Eliminates regression to mean
- Eliminates placebo effect
- Independent of rater bias if blindness maintained
- Basis for systematic reviews

for some purposes we may rather see efficacy and effectiveness trials as falling along a continuum. Efficacy trials, which usually precede effectiveness studies, refer to those conducted under more ideal, experimental conditions, while effectiveness trials are RCTs carried out in more routine clinical conditions.[28,40–42] Nevertheless, some important questions, for example the impact of clinical guidelines, may only be researchable in real world settings, and will therefore bypass the efficacy study stage.[43]

Cochrane has defined effectiveness, at the patient level, as assessing whether an intervention does more good than harm when provided under usual circumstances of health care practice.[28] At the level of service provision, Wells has defined effectiveness trials as those which 'duplicate as closely as possible the conditions in the target practice venues to which study results will be applied'.[44] The key differences between efficacy and effectiveness trials are shown in Table 7.6.12, although in practice the differences between these types of trial may not necessarily be as great as the differences between pharmacological, psychological, and service interventions

Table 7.6.9 Limitations and disadvantages of RCTs designs

1. Difficulties in choosing the unit or level of random allocation
   - Should allocation be made at the patient level, the clinician level, the clinical team/practice level, or the locality level?
2. Difficulties in achieving random allocation
   - Randomization not possible
   - Particular patient groups excluded
   - Self-exclusion because of non-consent
3. Difficulties in obtaining consent and in maintaining motivation
   - Consent may be inversely proportional to severity of condition
   - Consent may be refused because of patient treatment preferences
   - Retention with the trial may be affected by patient motivation
4. Difficulties in establishing and maintaining blindness
   - Degree of blinding of subjects
   - Degree of blinding of staff
   - Degree of blinding of raters
   - Deactive Hawthorne effect (the effect of being studied upon those being studied)
5. Difficulties related to the experimental conditions
   - Concurrent multiple interventions in health service research trials (without a single potentially active ingredient)
   - Interactions between treatment components
   - Consistency of control ('usual treatment') conditions
   - High attrition rates or loss to follow-up

   Large differences between conditions in which trials can take place and those of routine practice

Table 7.6.10 Situations when RCT designs are not applicable

1. Situations in which experimentation is inadequate
   - Poor generalizability—low external validity
   - Unrepresentative staff included
   - Atypical patients included
   - Treatments not standardized
2. Situations in which experimentation is impossible
   - Refusal of clinicians to take part
   - Ethical objections to the study
   - Political barriers
   - Legal objections
   - Contamination between experimental and control conditions
   - Scale of task—trials are required for too many treatments
3. Situations in which experimentation is inappropriate
   - Studies conducted to reduce the occurrence of events of very low frequency
   - Studies to prevent unwanted outcomes in the distant future
4. Experimentation unnecessary
   - When benefit/risk ratio is dramatic
   - When there is a small likelihood of confounders

(Reproduced from N. Black (1996), Why we need observational studies to evaluate the effectiveness of health care, *British Medical Journal*, **312** (7040), 1215–18, copyright 1997, BMJ Publishing Group Ltd.)

(such as the dissemination, and related barriers, of proven interventions).[45–47]

In planning effectiveness RCTs, seven sets of issues need to be carefully considered: (i) study question (e.g. is the study question expressed in an answerable way?), (ii) reference population (e.g. what is the reference group or subgroup to which the trial results should be generalized?), (iii) patient sample (e.g. how far does the sample reflect the target population?), (iv) study settings (e.g. how representative are the study settings of routine clinical sites?), (v) study interventions (e.g. is the study intervention manualized, acceptable to patients, and suitable for widespread use?), (vi) control condition (e.g. are the key characteristics of the control condition well described, and do they vary within and between sites?), and (vii) bias (e.g. attrition, blinding, concealment, consent, and

Table 7.6.11 Criteria to evaluate the quality of an RCT

Criteria
1. Is the hypothesis clearly defined?
2. Is the study population representative?
3. Was patient assignment randomized?
4. Were patients, practitioners, and assessors blind to the experimental intervention?
5. Were the groups similar at the start of the trial?
6. Were the groups treated equally apart from the experimental intervention?
7. Were all those who entered the trial accounted for at its conclusion?
8. Was this in the groups to which they were originally allocated?
9. Are all clinically important outcomes considered?
10. Whose perspective do they reflect?
11. Is the data analysis appropriate?
12. What is the size and precision of the treatment effect?
13. Do the likely benefits outweigh the harms and risks?
14. Is the conclusion supported by the results?

Table 7.6.12 Key differences between efficacy and effectiveness trials

	Efficacy trials	Effectiveness trials
Goal	To estimate efficacy and safety (if relevant) usually of a specific clinical intervention	To estimate relative benefits and risks of approved treatments, clinical interventions, programmes, or policies
When	Usually before an intervention is introduced	Post-implementation
Diagnosis	Diagnosis by structured interview	Clinical diagnosis or structured interview
Inclusion and exclusion criteria	Strict and multiple inclusion and exclusion criteria, typically excluding patients with comorbid physical and psychiatric disorders	Relatively few inclusion and exclusion criteria to optimize external validity of sample
Patient sample	Typically enrol highly motivated patients	Attempt to include more representative patients, including those who are ambivalent and who may not adhere to the allocated treatment regime
Sample size	At most a few hundred, more often less than 100	Often larger to enable smaller effect sizes to be identified in heterogeneous populations (e.g. in large simple trials with dichotomous outcomes)
Comparator	Placebo and/or single active comparator (for drug trials) Treatment as usual or active control (for psycho-social interventions)	One or more active comparators (for drug trials) Treatment as usual or active control (for psycho-social interventions)
Dosing	Fixed or flexible	Flexible dosing in clinically used range
Blinding	Triple-blind (i.e. patients, staff, and researchers blind), or double-blind	Double-blind (where possible), or single-blind
Duration	1–4 months	6 months or more
Research sites	Small number of experienced research sites	Dozens of routine treatment sites
Delivery of intervention	According to manual or protocol, not a focus of research	Fidelity to manual a key variable, and study may consider barriers to delivery of intervention in routine practice[47]
Research protocol	Strictly defined	Deliberately similar to usual practice
Adjunctive treatments	Not allowed or strictly limited	Allowed as in usual practice
Outcomes	Symptom rating scales and other clinical parameters	A single well-defined, clinically important outcome (for large simple trials) and multiple secondary outcomes, including safety and costs (for practical trials)

(Reproduced from S. Stroup, Practical clinical trials for schizophrenia, *Epidemiologia e Psichiatria Sociale*, **14**, 132–6, copyright 2005 Il Pensiero Scientifico Editore.)

contamination). These issues are shown in more detail in Tables 7.6.13 and 7.6.14.[48]

Quasi-experimental studies

The term 'quasi-experiment' was first used to refer to a situation in which the decision about whether an individual does or does not receive the intervention to be evaluated is not under the investigator's control.[49] Random allocation of patients is therefore not made, so selection bias may occur. In other respects, a quasi-experiment aims to apply the logic of randomized controlled trials to the study design, and the researcher tries to reduce this bias by making the study units, in the groups being compared, as alike as possible in terms of the most important characteristics. This approach is known as ***matching***. Characteristics chosen for matching are those expected to influence the outcome (i.e. confounding factor). Therefore, the overriding aim of matching is to reduce the contribution made by the matched variables to the selection bias, although this method is inferior to randomization in that it cannot reduce the selection bias from all other variables.

There are two main approaches to matching. ***Paired matching*** consists of selecting individuals for the comparison group (or groups) who have closely similar characteristics to those included in the experimental group, for example in terms of age, gender, and occupation. This form of prestratification will need to be taken into account at the data analysis stage. A less rigorous variant is *group matching*, which only ensures that there are similar overall proportions of people, for both the experimental and comparison groups, in the various age bands, occupational groups, or other predefined strata used for the variables chosen for matching.

Non-experimental descriptive studies

The next type of research design included in the hierarchy of evidence is non-experimental descriptive studies. For the sake of clarity and brevity we shall distinguish two types of descriptive study: structured clinical practice and everyday unstructured clinical practice. An example of a descriptive evaluation design is the South Verona Outcome Study.[50,51] This is a prospective study, which aims to

Table 7.6.13 Criteria to plan effectiveness randomized controlled trials[48]

1. Study question
   - Who defines the aim of the study?
   - What process is used to identify the question addressed?
   - Is the study question expressed in an answerable way? (as a clear hypothesis)
   - Prior evidence of intervention effect size
   - Is the answer to this question really unknown?
   - Why is this question important now?
   - Is there initial evidence from efficacy trials or effectiveness studies? (observational or trials)
   - What is the public health importance of the policy or practice question addressed?
   - What is the clinical necessity of the question?
   - Sample size and statistical power for primary/secondary aims and related hypotheses
2. Reference population
   - What is the reference group (or subgroup) to which the trial results should be generalized?
   - What are their socio-demographic and clinical characteristics?
   - What are the ethnic and cultural characteristics of the target group?
   - What is resource level in this population?
   - What is the nature and standard and coverage of health and social care?
   - At what time point is population identified?
3. Patient sample
   - What are their socio-demographic, and clinical characteristics?
   - What are the inclusion criteria?
   - Not invited to participate rate
   - Non-participation rate
   - Patient preferences
   - What are the exclusion criteria?
   - How far does the sample reflect the target population?
   - What level of heterogeneity is there?
   - Selection of incident or prevalent cases (true incidence/prevalence or treated incidence/prevalence)
   - What are the rates of adherence and non-adherence to treatment as recommended?
4. Study settings
   - Characteristics and representativeness of professional staff
   - Levels of resources available
   - Research oriented culture
   - Staff morale and sustainability of intervention
   - Incentives for research collaboration
   - Opportunities for data linkage
   - Centre/professional non-participation
5. Study intervention
   - Is intervention acceptable?
   - Total time needed to deliver intervention
   - Frequency of interventions
   - Simplicity/complexity of the intervention
   - Single/multi-component intervention
   - Is intervention manualized?
   - Do usual professional staff deliver the intervention during the study?
   - Can treatment process be measured? (fidelity)
   - Degree of fit/feasibility for current practice
   - Exit strategy, who pays after the end of study
6. Control condition
   - Treatment as usual or specific control
   - Acceptability to patients of control condition
   - Cost and feasibility of control condition
   - Variation between control condition within and between sites (fidelity)
   - Are the key characteristics of the control condition well described?
7. Bias
   - Does contamination take place?
   - Degree of blinding
   - Choice of primary and secondary outcomes
   - Perspectives prioritized in outcome choice
   - Time(s) at which outcomes measured
   - Total length of follow-up and late effects
   - Sources of outcome data
   - Respondent burden
   - Consent rate
   - Recruitment rate
   - Attrition/drop-out and follow-up rates

(Reproduced from Tansella, M. Thornicroft, G., Barbui, C. *et al.* Seven criteria for improving effectiveness trials in psychiatry. *Psychological Medicine*, **36**(5) 711–200, copyright 2006, Cambridge University Press.)

assess the outcome of mental health care. Data from this study have been analysed using a multidimensional perspective. Among 354 patients followed up after 6 years of treatment in routine clinical settings the study revealed a complex pattern of emerging and disappearing clinical and patterns of exacerbation and remissions, with both changing frequently over time, but changes in both clinicial and social domains were not associated with diagnosis.[52]

Conclusions

Over the course of the next decade we expect that the following key trends will be of paramount importance. In relation to the focus of research, in many countries a degree of contestability may well develop, in which those who have traditionally identified questions to be addressed by research (investigators) will be challenged by research funders (such as governments and charities) and by the intended beneficiaries of the research (people with mental illness and their family members) to set the research agenda.[53] We can expect governments to direct their research investment towards policy challenges, such as barriers to the implementation of evidence-based practice.[54–56] This may well include the commissioning of research not just on new treatments and services, but also to evaluate already or even long-established models of care. For example, there is relatively little mental health service evaluation about: outpatient services (clinics), inpatient services, or forensic service provision.[57,58] In future, it will be important to evaluate *post hoc* current but unproven service configurations, particularly those that are widespread and expensive, as well as innovative interventions. This will necessitate providing sufficient long-term funding for health service research.

In terms of study design, we anticipate that there will be a relative growth of effectiveness studies, especially RCTs, which attempt to balance internal and external validity.[48] These will more often than in the past specify the precise nature of the control condition, use representative patient samples, and standardized outcomes measures. Less common study designs, such as cluster and preference RCTs will be necessary to tackle complex interventions.[59,60] The recent trend to more often include qualitative assessments

within RCTs we expect to accelerate, for example to identity the acceptability of interventions to patients, and to identify the active ingredients (and barriers) to treatment effectiveness.[61–65]

How will the conduct of evaluations of mental health services evolve in the coming years? We can identify a trend for study interventions to be increasingly often manualized. At the same time the patient populations treated will be more often similar to those treated in routine clinical practice. As a consequence, more attention will need to be paid to ways to incentivize clinical staff to participate in research.

Although traditionally research scientists have seen the dissemination of research findings largely in terms of publications in scientific journals,[66] it is likely that research funders will increasingly encourage or even insist upon using effective and target-specific communication methods to reach key audiences with the results of research and their implications, including non-traditional methods such as social marketing.[67] This will be intended to alter the behaviour of service planners, commissioners, practitioners, and service users, so that the results of research do influence the behaviour of these key groups in terms of service decisions, so that they increasingly reflect the evidence both of effective interventions and where the evidence shows lack of effect then to stop ineffective practices and services. Such behaviour change (and an established evidence base for this) may include methods as social marketing, as well as carefully combined interventions such as those used in case management in the treatment of depression.[68]

Table 7.6.14 Key challenges to the evaluation of mental health services

Focus of research
◆ Clarifying who is defining research questions
◆ Including service users and family members in setting research questions
◆ Asking clear research questions to answer important clinical challenges
◆ Evaluating already established as well new services
◆ Providing sufficient funding for long-term health service research
Study design
◆ Balancing internal and external validity of mental health service evaluation
◆ Moving from efficacy to effectiveness trials
◆ Specifying the precise nature of the control condition
◆ Using representative patient samples
◆ Using standardized outcome measures
◆ Combining qualitative and quantitative information
Conduct of research studies
◆ Manualizing the interventions to be evaluated
◆ Specifying the key characteristics of the patient groups to be treated
◆ Incentivizing clinical staff to participate in research
Data analysis and interpretation of findings
◆ Identifying the active ingredients of effective interventions
Dissemination of research findings
◆ Using effective and target-specific communication methods to reach key audiences with the results of research and their implications
Implementation of effective interventions
◆ Implementing the results of evaluation when the evidence is strong enough and decommissioning ineffective practice

Further information

Thornicroft, G. and Tansella, M. (1999). *The mental health matrix: a manual to improve services*. Cambridge University Press, Cambridge.

Thornicroft, G. and Szmukler, G. (2001). *Textbook of community psychiatry*. Oxford University Press, Oxford.

Tansella, M. and Thornicroft, G. (2001). *Mental health outcome measures* (2nd edn). Gaskell, Royal College of Psychiatrists, London.

Thornicroft, G. (2001). *Measuring mental health needs* (2nd edn). Gaskell, Royal College of Psychiatrist, London.

Fulop, N. and Allen, P. (2002). *Studying the organization and the delivery of the health services: research methods*. Routledge, London.

Thornicroft, G. and Tansella, G. (2004). The components of a modern mental health service: a pragmatic balance of community and hospital care. *The British Journal of Psychiatry*, **185**, 283–90.

Thornicroft, G., Becker, T., Knapp, M., *et al.* (2006). International outcome measures in mental health. Quality of life, needs, service satisfaction, costs and impact on carers. Gaskell, Royal College of Psychiatrists, London.

Slade, M. and Priebe, S. (2006). *Choosing methods in mental health research*. Routledge, London.

Knapp, M.J., McDaid, D., Mossialos, E., *et al.* (2007). *Mental health policy and practice across Europe*. Open University Press, Buckingham.

References

1. Campbell, N.C., Murray, E., Darbyshire, J., *et al.* (2007). Designing and evaluating complex interventions to improve health care. *British Medical Journal*, **334**(7591), 455–9.
2. Campbell, M., Fitzpatrick, R., Haines, A., *et al.* (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, **321**, 694–6.
3. Soanes, C. and Stevenson, A. (2003). *Concise Oxford English dictionary* (11th edn). Oxford University Press, Oxford.
4. Thornicroft, G. and Tansella, M. (1999). *The mental health matrix: a manual to improve services*. Cambridge University Press, Cambridge.
5. Thornicroft, G. and Tansella, M. (2007). *Better mental health care*. Cambridge University Press, Cambridge.
6. Sartorius, N. (1997). Evaluating mental health services. A world perspective. *Epidemiologia e Psichiatria Sociale*, **6**(Suppl. 1), 239–45.
7. Johnson, S., Nolan, F., Hoult, J., *et al.* (2005). Outcomes of crises before and after introduction of a crisis resolution team. *The British Journal of Psychiatry*, **187**, 68–75.
8. Killaspy, H., Bebbington, P., Blizard, R., *et al.* (2006). The REACT study: randomised evaluation of assertive community treatment in north London. *British Medical Journal*, **332**(7545), 815–20.
9. Glover, G., Arts, G., and Babu, K.S. (2006). Crisis resolution/home treatment teams and psychiatric admission rates in England. *The British Journal of Psychiatry*, **189**(5), 441–5.
10. Tansella, M. and Thornicroft, G. (eds.) (2001). *Mental health outcome measures*. Royal College of Psychiatrists, Gaskell, London.
11. Thornicroft, G., Becker, T., Knapp, M., *et al.* (2006). International outcome measures in mental health. Quality of life, needs, service satisfaction, costs and impact on carers. Gaskell, Royal College of Psychiatrists, London.
12. Schene, A., Tessler, R.C., Gamache, G.M., *et al.* (2001). Measuring family or care giver burden in severe mental illness: the instruments. In *Mental health outcome measures* (2nd edn) (eds. M. Tansella and G. Thornicroft), pp. 48–71. Royal College of Psychiatrists, Gaskell, London.
13. Joyce, J., Leese, M., and Szmukler, G. (2000). The experience of caregiving inventory: further evidence. *Social Psychiatry and Psychiatric Epidemiology*, **35**(4), 185–9.
14. Thornicroft, G. and Tansella, M. (2004). The components of a modern mental health service: a pragmatic balance of community and hospital care. *The British Journal of Psychiatry*, **185**, 283–90.

15. Szmukler, G., Kuipers, E., Joyce, J., *et al.* (2003). An exploratory randomised controlled trial of a support programme for carers of patients with a psychosis. *Social Psychiatry and Psychiatric Epidemiology*, **38**(8), 411–8.
16. Ruggeri, M. (2001). Measuring satisfaction with psychiatric services: towards a multi-dimensional, multi-axial assessment of outcome. In *Mental health outcome measures* (2nd edn) (eds. M. Tansella and G. Thornicroft), pp. 34–47. Royal College of Psychiatrists, Gaskell, London.
17. Ruggeri, M., Dall'Agnola, R., Agostini, C., *et al.* (1994). Acceptability, sensitivity and content validity of the VECS and VSSS in measuring expectations and satisfaction in psychiatric patients and their relatives. *Social Psychiatry and Psychiatric Epidemiology*, **29**(6), 265–76.
18. Lehman, A. (2001). Measures of quality of life for people with severe mental disorders. In *Mental health outcome measures* (2nd edn) (eds. M. Tansella and G. Thornicroft), pp. 72–92. Royal College of Psychiatrists, Gaskell, London.
19. Ware, J. and Sherbourn, C. (1992). The MOS, 36 item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, **30**, 473–83.
20. Wiersma, D. (2001). Measuring social disabilities in mental health. In *Mental health outcomes measures* (2nd edn) (eds. M. Tansella and G. Thornicroft), pp. 118–32. Royal College of Psychiatrists, Gaskell, London.
21. Thornicroft, G. (2001). *Measuring mental health needs* (2nd edn). Royal College of Psychiatrists, Gaskell, London.
22. Slade, M., Thornicroft, G., Loftus, L., *et al.* (1999). *CAN: the Camberwell Assessment of Need*. Gaskell, Royal College of Psychiatrists, London.
23. Lasalvia, A., Bonetto, C., Malchiodi, F., *et al.* (2005). Listening to patients' needs to improve their subjective quality of life. *Psychological Medicine*, **35**(11), 1655–65.
24. Slade, M., Leese, M., Ruggeri, M., *et al.* (2004). Does meeting needs improve quality of life? *Psychotherapy and Psychosomatics*, **73**(3), 183–9.
25. Geddes, J.R. and Harrison, P.J. (1997). Closing the gap between research and practice. *The British Journal of Psychiatry*, **171**, 220–5.
26. L'Abbe, K.A., Detsky, A.S., and O'Rourke, K. (1987). Meta-analysis in clinical research. *Annals of Internal Medicine*, **107**(2), 224–33.
27. Sackett, D.L., Rosenberg, W.M., Gray, J.A., *et al.* (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, **312**(7023), 71–2.
28. Cochrane, A. (1972). *Effectiveness and efficiency: random reflections on health services*. Nuffield Provincial Hospitals Trust.
29. Tyrer, P., Coid, J., Simmonds, S., *et al.* (2000). Community mental health teams (CMHTs) for people with severe mental illnesses and disordered personality. *Cochrane Database of Systematic Reviews*, (2), CD000270.
30. Simmonds, S., Coid, J., Joseph, P., *et al.* (2001). Community mental health team management in severe mental illness: a systematic review. *The British Journal of Psychiatry*, **178**, 497–502.
31. Korn, E.L. and Baumrind, S. (1991). Randomised clinical trials with clinician-preferred treatment. *Lancet*, **337**(8734), 149–52.
32. Barker, D. and Rose, G. (1979). *Epidemiology in medical practice* (2nd edn). Churchill Livingstone, London.
33. Everitt, B. and Wessely, S. (2004). *Clinical trials in psychiatry*. Oxford University Press, Oxford.
34. Black, N. (1996). Why we need observational studies to evaluate the effectiveness of health care. *British Medical Journal*, **312**(7040), 1215–18.
35. Marriott, S. and Palmer, C. (1996). Clinical practice guidelines: on what evidence is our clinical practice based? *Psychiatric Bulletin*, **20**, 363–6.
36. Macfarlane, W., Dushay, R., Stastny, P., *et al.* (1996). Comparison of two levels of family-aided assertive community treatment. *Psychiatric Services*, **47**, 744–50.
37. Schwartz, D. and Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutic trials. *Journal of Chronic Diseases*, **20**(20), 637–48.
38. Peto, R., Collins, R., and Gray, R. (1993). Large scale randomised evidence. *Annals of the New York Academy of Science*, **703**, 314–40.
39. Oliver, S. (1997). Exploring lay perspectives on questions of effectiveness. In *Non random reflections on health services research* (eds. A. Maynard and I. Chalmers), pp. 272–91. British Medical Journal Publications, London.
40. Haynes, B. (1999). Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *British Medical Journal*, **319**(7211), 652–3.
41. Lilienfeld, A. (1982). The Fielding H. Garrison lecture: ceteris paribus: the evolution of the clinical trial. *Bulletin of the History of Medicine*, **56**, 1–18.
42. Pocock, S. (1983). *Clinical trials: a practical approach*. Wiley, London.
43. Andrews, G. (1999). Randomised controlled trials in psychiatry: important but poorly accepted. *British Medical Journal*, **319**(7209), 562–4.
44. Wells, K.B. (1999). Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. *The American Journal of Psychiatry*, **156**(1), 5–10.
45. Proudfoot, J., Goldberg, D., Mann, A., *et al.* (2003). Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. *Psychological Medicine*, **33**(2), 217–27.
46. Wells, K.B., Sherbourne, C., Schoenbaum, M., *et al.* (2000). Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *The Journal of the American Medical Association*, **283**(2), 212–20.
47. Campbell, M., Fitzpatrick, R., Haines, A., *et al.* (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, **321**(7262), 694–6.
48. Tansella, M., Thornicroft, G., Barbui, C., *et al.* (2006). Seven criteria for improving effectiveness trials in psychiatry. *Psychological Medicine*, **36**(5), 711–20.
49. Campbell, D.T. and Stanley, J.C. (1996). *Experimental and quasi-experimental designs for research*. Rand-McNally, Chicago, IL.
50. Ruggeri, M., Biggeri, A., Rucci, P., *et al.* (1998). Multivariate analysis of outcome of mental health care using graphical chain models. The south-Verona outcome project 1. *Psychological Medicine*, **28**(6), 1421–31.
51. Lasalvia, A. and Ruggeri, M. (2007). Multidimensional outcomes in 'real world' mental health services: follow-up findings from the south Verona project. *Acta Psychiatrica Scandinavica*, (Suppl.).
52. Lasalvia, A., Bonetto, C., Cristofalo, D., *et al.* (2007). Predicting clinical and social outcome of patients attending 'real world' mental health services: a 6 year multi-wave follow-up. *Acta Psychiatrica Scandinavica*, **116**, (s437), 16–30.
53. Chamberlin, J. (2005). User/consumer involvement in mental health service delivery. *Epidemiologia e Psichiatria Sociale*, **14**(1), 10–4.
54. Magnabosco, J.L. (2006). Innovations in mental health services implementation: a report on state-level data from the U.S. Evidence-based practices project. *Implemention Science*, **1**, 13.
55. Drake, R.E., Becker, D.R., Goldman, H.H., *et al.* (2006). Best practices: the Johnson & Johnson—Dartmouth community mental health program: disseminating evidence-based practice. *Psychiatric Services*, **57**(3), 302–4.
56. Aarons, G.A. and Sawitzky, A.C. (2006). Organizational culture and climate and mental health provider attitudes toward evidence-based practice. *Psychological Services*, **3**(1), 61–72.
57. Szmukler, G. and Holloway, F. (2001). In-patient treatment. In *Textbook of community psychiatry* (eds. G. Thornicroft and G. Szmukler), pp. 321–37. Oxford University Press, Oxford.
58. Becker, T. (2001). Out-patient psychiatric services. In *Textbook of community psychiatry* (eds. G. Thornicroft and G. Szmukler), pp. 277–82. Oxford University Press, Oxford.

59. Medical Research Council. (2002). *Cluster randomised trials: methodological and ethical considerations*. Medical Research Council, London.
60. Howard, L. and Thornicroft, G. (2006). Patient preference randomised controlled trials in mental health research. *The British Journal of Psychiatry*, **188**, 303–4.
61. Lester, H., Tritter, J.Q., and Sorohan, H. (2005). Patients' and health professionals' views on primary care for people with serious mental illness: focus group study. *British Medical Journal*, **330**(7500), 1122.
62. Pope, C., Mays, N., and Popay, J. (2006). How can we synthesize qualitative and quantitative evidence for healthcare policy-makers and managers? *Healthcare Management Forum*, **19**(1), 27–31.
63. Mays, N., Pope, C., and Popay, J. (2005). Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *Journal of Health Services Research & Policy*, **10**(Suppl. 1), 6–20.
64. Pope, C., Ziebland, S., and Mays, N. (2000). Qualitative research in health care. Analysing qualitative data. *British Medical Journal*, **320**(7227), 114–16.
65. Mays, N. and Pope, C. (2000). Qualitative research in health care. Assessing quality in qualitative research. *British Medical Journal*, **320**(7226), 50–2.
66. Lewison, G., Thornicroft, G., Szmukler, G., *et al*. (2007). The fair assessment of the merits of psychiatric research. *The British Journal of Psychiatry*, **190**, 314–18.
67. Kotler, P., Roberto, E.L., and Lee, N. (2002). *Social marketing: improving the quality of life*. Sage, New York.
68. Wells, K., Miranda, J., Bruce, M.L., *et al*. (2004). Bridging community intervention and mental health services research. *The American Journal of Psychiatry*, **161**(6), 955–63.
69. Thornicroft, G. (2006). *Shunned: discrimination against people with mental illness*. Oxford University Press, Oxford.
70. Pinfold, V., Toulmin, H., Thornicroft, G., *et al*. (2003). Reducing psychiatric stigma and discrimination: evaluation of educational interventions in UK secondary schools. *The British Journal of Psychiatry*, **182**, 342–6.
71. Sackett, D.L. and Wennberg, J.E. (1997). Choosing the best research design for each question. *British Medical Journal*, **315**(7123), 1636.
72. Stroup, S. (2005). Practical clinical trials for schizophrenia. *Epidemiologia e Psichiatria Sociale*, **14**, 132–6.

7.7

Economic analysis of mental health services

Martin Knapp and Dan Chisholm

Introduction

Economics is concerned with the use and distribution of resources within a society, and how different ways of allocating resources impact on the well-being of individuals. Economics enters the health sphere because resources available to meet societal needs or demands are finite, meaning that choices have to be made regarding how best to allocate them (typically to generate the greatest possible level of population health). Economics provides an explicit framework for thinking through ways of allocating resources.

Resource allocation decisions in mental health are complicated by the fact that disorders are common, debilitating, and often long-lasting. Epidemiological research has demonstrated the considerable burden that mental disorders impose because of their prevalence, chronicity, and severity: globally, more than 10 per cent of lost years of healthy life and over 30 per cent of all years lived with disability are attributable to mental disorders.[1] Low rates of recognition and effective treatment compound the problem, particularly in poor countries.

However, disease burden is not in itself sufficient as a justification or mechanism for resource allocation or priority-setting. A disorder can place considerable burden on a population but if appropriate strategies to reduce this burden are absent or extremely expensive in relation to the health gains achieved, large-scale investment would be considered misplaced. The reason is that scarce resources could be more efficiently channelled to other burdensome conditions for which cost-effective responses *were* available. For priority-setting and resource allocation, it is necessary to ask what amount of burden from a disorder can be avoided by using evidence-based interventions, and at what relative cost of implementation in the target population.

Cost and cost-effectiveness considerations enter into health care reform processes, priority-setting exercises within and across health programmes, and regulatory decisions concerning drug approval or pricing. Two broad levels of economic analysis can be distinguished: macro and micro.

Economic analyses at macro level: the mental health system

Macro-level economic analyses are concerned with how health systems function and what they achieve. What, for example, are the motivations and behaviour of key 'stakeholders', with what implications for access, quality, and costs? What roles do economic forces play and can they be shaped to improve health outcomes and cost-effectiveness? Do different organizational or financial arrangements produce different resource configurations? For example, do markets achieve fairer or more efficient allocations than state bureaucracies? Through their macro analyses, economists can contribute to a better understanding of how health systems can improve utilization of resources (for example, see Box 7.7.1).

While improved psychological well-being in the population is likely to represent the primary goal of the mental health system, there are other (social) goals that could also feature prominently, including quality improvements in service provision, and financial (as well as human rights) protection for people with mental health problems.[2] Meeting these goals is achieved via a number of key health system functions, including resource generation, their allocation via appropriate modes of financing, actual provision of services, and overall stewardship and evaluation of these various functions.[3] Economic analysis contributes to policy formation relating to each of these functions.[4–6]

There are barriers to implementation of evidence-based mental health care in even generously resourced health systems. An insidious barrier is resource insufficiency: mental health services are often grossly under-funded in comparison to both needs and the

Box 7.7.1 The relevance of a health systems perspective

The need for a systems approach to mental health policy and planning is made apparent from a simple illustration: cheap, effective drugs exist for key neuropsychiatric disorders, including tricyclic antidepressants, conventional neuroleptics, and antiepileptic drugs, which are affordable even to resource-poor countries. The availability and prescription of these drugs to those in need, however, are determined by the extent to which such drugs have been distributed and by the ability of health care providers to detect and appropriately treat the underlying condition. Access to and use of such medications may further be hampered by the private cost of seeking and receiving health care, particularly if it is out-of pocket. User fees, provider incentives, and clinical practice are in turn influenced by the availability of national legislation, regulation, and treatment guidelines.

cost-effectiveness of interventions to meet them. This is a major issue for countries where the proportion of national income devoted to health care is low, or where the proportion of the health budget allocated to mental health is minimal. With limited funds it is difficult to build any kind of service system, because it is difficult to recruit, train, and retain skilled staff.

Even when resources are committed, available services might be poorly distributed, available at the wrong place or time relative to the distribution of needs. They may only be delivered by specialist clinics or concentrated in big cities, or affordable only by wealthier individuals. Improvements to practice take time to work through to improved health outcomes, cost-effectiveness gains or fairer access, even when suitable professionals can be recruited or new facilities opened. Decision makers must think long-term, for the immediate consequences of many interventions could be modest but longer-term benefits immense.[7]

A more general difficulty is that available services do not match what is needed or preferred. Indeed, there may be scant information on population or individual needs, and patients may have few opportunities to participate in treatment decisions. Another problem could be poor coordination of services because of professional rivalry, stultifying bureaucracy or 'silo budgeting' (resources held in one agency's 'silo' cannot be allocated to other uses).

Economic analyses at micro level: cost-effectiveness

The most frequently posed micro questions relate to the cost-effectiveness of interventions, such as an emphasis on community-based care, the use of new drugs, or the development of secure accommodation. For example, consider what happens following development of a new treatment (say a new medication for schizophrenia). Decision makers want answers to two questions when considering whether to use or recommend this drug. The first is the clinical question: is it effective in alleviating psychotic symptoms and generally improving health-related quality of life? If the answer to the clinical question is 'yes', then there is a second question: is it cost-effective? That is, does the drug achieve the improved outcomes at a cost that is worth paying? The meaning of 'worth' is far from straightforward to establish and laden with controversy.

These two questions sit at the heart of economic evaluation: outcomes must be assessed (and compared between different treatments) and the (relative) costs of achieving them must be examined. Looking only at costs is *not* an economic evaluation.

There are different variants of economic evaluation. They share a common approach to the conceptualization, definition, and measurement of costs, but adopt different approaches when considering and measuring outcomes, primarily because they seek to answer slightly different questions. We set out these differences by discussing the questions a study might address, measuring costs and outcomes, making trade-offs between them, and utility and benefit measurement.

Question and perspective

The choice of evaluative approach depends on the question to be addressed. If the question is essentially clinical—what is the most appropriate treatment for someone with particular needs in particular circumstances—information is needed on the comparative costs of alternative treatments and comparative outcomes measured in terms of symptom alleviation, improved functioning, and so on. A cost-effectiveness analysis would be appropriate (see 'Effectiveness measurement' below).

If, to take a broader stance, the question is whether to treat depression rather than spending the funds elsewhere in the health system, then decision makers need to know the costs, but now need an outcome measure that uses a common metric across different health domains. The most common such metric is 'utility' and a cost-utility analysis would be undertaken (see 'Utility measurement' below).

To widen the perspective further, if the question is whether to increase expenditure in the health system or in (say) improving transport or launching a new environmental policy, then an evaluation needs to ask about the comparative costs and impacts of the different options, where 'impact' will need to be measured in a common unit across all public policy areas. The usual choice of broad measure is monetary, leading to cost–benefit analysis (see 'Benefit measurement' below).

The question to be addressed thus influences the type of evaluation needed, but the choices are not mutually exclusive: a single study can support more than one approach if the right measures are used. The broader the question, the lower the likelihood that the outcome measure will be sensitive to the particular circumstances of a specific disorder such as depression, but the greater the usefulness in terms of resource allocation decisions.

Linked to specification of the question to be addressed is the *perspective* of a study. Is the evaluation needed to help resource allocation within a particular agency (such as primary care clinic), or a particular system (such as the health care system), or the whole society? The perspective will determine the breadth of both cost and outcome measurement.

Cost measurement

Some costs are directly associated with a disorder or its treatment, such as the money spent on medications and services used by patients, and some are more indirect, measuring lost productivity because ill-health can disrupt someone's employment pattern or the social cost of unpaid care provided by families. How broadly the costs are measured will depend upon the purpose of the study.

In carrying out evaluations in practice, economists need data on service use patterns by patients. This information might come from organizational 'billing' systems (recording amounts transferred between purchasers and providers for services used), or from routine information systems that record service contacts, or from research instruments that specifically collect data on service use patterns through interviews with patients, caregivers, or service professionals. One widely used instrument is the Client Service Receipt Inventory.[8]

The next task is to attach unit cost estimates to these service use data. In England, there is an excellent annual compendium of health and social care unit costs, which provides just such figures.[9] In other countries, it might be necessary to estimate unit costs anew. A range of data sources could be used, including government statistics, health system expenditure figures, and specific facility or organization accounts. The main cost categories to be quantified would be:

- salaries of staff employed in patient treatment and care
- facility operating costs (e.g. cleaning, catering)

- overhead costs (e.g. personnel, finance)
- capital costs for buildings and durable equipment

Effectiveness measurement

The most intuitive mode of economic evaluation is cost-effectiveness analysis (CEA): it measures costs as set out above, and outcomes along the dimensions that would be recognized by clinicians and used in clinical studies (changes in symptoms, behaviour, functioning, and so on). A CEA can help decision makers choose between interventions aimed at specific health needs. A cost-effectiveness analysis looks at a single outcome dimension—such as change in symptoms—and computes and compares the difference in costs between two treatments and the difference in this (primary) outcome. If one treatment is both more effective and less costly than another, then it would clearly be the more cost-effective of the two. But if it is more effective and *more* costly then a trade-off is needed (see below).

Often the economist will compute cost differences and a range of effectiveness differences (one for each outcome dimension)—an approach sometimes called cost-consequences analysis—which has the advantage of breadth but poses a challenge if one outcome is better and another worse for a particular treatment. It is then not always obvious which treatment is to be preferred, and the decision maker must weigh up the strength of evidence.

Making trade-offs

If an evaluation finds a new intervention to be both more effective but simultaneously more expensive than an older intervention, which is the more cost-effective of the two? A trade-off must be made between the better outcomes and the higher costs necessary to achieve them.

The classical way of determining this trade-off has been via the derivation of a incremental cost-effectiveness ratio (CER), which divides the extra cost associated with a new intervention by its additional effect (see Box 7.7.2).

More recently, health economists have developed the 'net benefit approach' to explicate the nature of the trade-off (see Box 7.7.3 for an example). It is commonly seen today in the construction of *cost-effectiveness acceptability curves* (CEACs). These curves show the probability that an intervention will be cost-effective for each of a number of pre-specified or implicit valuations of an outcome improvement by the decision maker.

Utility measurement

One way to overcome the potential problem of different outcome dimensions pointing in different directions is to employ a single, over-arching measure. A preference-weighted, health-related quality of life measure could be used. The value of the quality of life improvement is gauged in units of 'utility', usually expressed by a combined index of the mortality and quality of life effects of an intervention. The best known such index is the Quality Adjusted Life Year (QALY).

A cost-utility analysis (CUA) measures the outcome difference between two interventions in terms of QALY gain, and compares this with the difference in costs. CUAs have a number of attractions, including a unidimensional, generic outcome measure that allows comparisons across diagnostic groups, based on an explicit methodology for weighting preferences and valuing health states. But the utility measure may be too reductionist and insufficiently sensitive to changes expected in a particular clinical area such as depression treatment.[11] Nevertheless, cost-utility analyses produce estimates of cost-per-QALY gain from one therapy over another, which can then inform health care resource allocation decisions, such as by the National Institute for Health and Clinical Excellence (NICE) in England and Wales.

Box 7.7.2 Calculation of the incremental cost-effectiveness ratio

For example, evaluation of a new antidepressant may show that the cost per treated case of major depression is an extra £500 per year but also results in a lower average symptom score over this period (such as a ten point reduction on the Beck Depression Inventory); the resulting incremental CER would therefore be £50 [£500/10], meaning that each additional unit improvement on the BDI cost £50.

Box 7.7.3 Example of the net benefit approach in mental health services research

An example of the use of cost-effectiveness acceptability curves comes from a study of computer-delivered cognitive behavioural therapy (CCBT) for anxiety and depression.[10] CCBT was more expensive in health service terms than standard primary care services, but more effective in reducing symptoms. The fitted CEACs showed that, even if the value placed by society on a unit reduction in the Beck Depression Inventory (the primary clinical measure used in the trial) was as little as £40, there was an 81 per cent probability that CCBT would be viewed as cost-effective. Similarly, assigning a societal value of just £5 to each additional depression-free day would result in an 80 per cent probability that CCBT would be cost-effective. The CEAC makes transparent the trade-offs faced by decision makers.

Benefit measurement

Cost–benefit analysis asks whether the benefits of a treatment or policy exceed the costs, helping decision makers to allocate resources across a wide area, for example comparing health care with housing, education, or defence. All costs and outcomes (benefits) are valued in the same (monetary) units. If benefits exceed costs, the evaluation would provide support for the intervention or programme, and vice versa. With two or more alternatives, the intervention with the greatest net benefit would be deemed the most efficient. Cost–benefit analyses are thus intrinsically attractive, but conducting them is especially problematic because of the difficulties associated with attaching monetary values to health outcomes, and especially mental health outcomes. Methodological advances in health economics offer ways to obtain direct valuations of health outcomes by patients, families, or others,[12] but they will not be easy to apply in mental health contexts.

Design issues

As in clinical evaluation, an important consideration for the review, assessment, and interpretation of economic evidence is research

design. Generally speaking, the ideal type of study upon which to base decisions on cost-effectiveness and resource allocation is the one conducted prospectively with two (or more) appropriately sized randomly allocated groups of patients, for whom all conceivable costs and outcomes are measured appropriately, including a comparable measure of outcome (monetized benefits or, more realistically, a utility metric).

Looking across the mental health field, the accumulation of new cost-effectiveness evidence has been uneven, tending to be greater in diagnostic areas where new classes of medication have been launched: the pharmaceutical industry looks to economic evidence to support its marketing. At the same time, health care funding and delivery bodies also want their own independent evidence on new therapies. Consequently, a lot of economic studies of depression followed the licensing of the early selective serotonin-reuptake inhibitors (SSRIs) and later antidepressants with other mechanisms of action. Similarly, the arrival of the atypical antipsychotics and the cholinesterase inhibitors for Alzheimer's disease stimulated a lot of economics research.

We cannot cover all mental health areas here. Instead, in the next three sections, we look at areas where there has been some interesting activity:

- cost-utility analysis of depression treatment in primary care
- cost-effectiveness of interventions for child and adolescent mental health problems
- sectoral cost-effectiveness analysis of mental health interventions in developing countries.

Cost-utility analysis of depression treatment in primary care

Ten years ago, examples of the application of the cost-utility approach to mental health were hard to find.[11] Since then, there has been an increasing use of the so-called cost-per-QALY approach in mental health evaluations, following recommendations for such analyses by regulatory bodies in Australia, Canada, the United Kingdom, and the United States. In the field of depression, for example, cost-utility analyses have now been carried out for: screening in primary care[13] (annual and periodic screening cost more than $50 000 per QALY, one-off screening below this threshold); newer versus older antidepressant drugs[14]; maintenance treatment for recurrent depression[15,16]; guideline-concordant primary care treatment for women[17]; primary care practice-initiated quality improvement programmes[18]; computerized cognitive behavioural therapy[10]; and ECT versus transcranial magnetic stimulation for 'treatment-resistent' depression.[19] Many of the cost-utility analyses carried out to date employ secondary data and modelling techniques to estimate costs and effects, others have constructed cost-per-QALY estimates alongside clinical trials.

An example of a cost-utility analysis using modelling is that by Revicki *et al.*[14] who compared treatment for major depression with (a) newer antidepressants (nefazodone and fluoxetine) (b) tricyclics (imipramine) and, for treatment failures, (c) a step approach involving initial treatment with imipramine followed by nefazodone. A decision analysis model was developed to simulate the clinical management pathways and pattern of recurrences of major depression for these alternative treatment strategies to estimate lifetime medical costs and health outcomes (expressed as QALYs). There were only minor differences in costs and QALYs between nefazodone and fluoxetine, and both these newer antidepressants were estimated to be cost-effective compared to imipramine treatment and the imipramine step approach. The ratios of cost to QALYs gained for these newer antidepressants were deemed to be sufficiently low (below $20 000 per QALY gained) to merit adoption of these treatments in the health system. For example, the extra lifetime cost of nefazodone over imipramine ($1321) resulted in 0.32 added QALYs, giving a ratio of $4065 per QALY gained. Since decision models and their findings are only as good as their underlying assumptions and the quality of the data used to estimate key model parameters, extensive sensitivity analyses were conducted, but these did not alter the conclusions. However, the results did not include indirect costs such as changes in work productivity—important for a societal perspective—and are not readily generalizable to groups other than the targeted population (in this study, 30-year old women with one previous depressive episode).

An example of the empirically based generation of cost-per-QALY information is the randomized controlled trial of practice-initiated quality improvement (QI) for depression,[18] which involved group-level randomization of 46 clinics in six community-based US managed care organizations, either to medication or psychotherapy quality improvement programme (in addition to training and enhanced educational resources). Two QALY measures were derived, one from the Short-From, 12-Item Health Survey (SF-12) plus a standard gamble utility weighting exercise among a local convenience sample, the other with reference to estimated time spent depressed plus values from the literature for lost utility due to depression. Relative to usual care, average health care costs increased by $400–500 per treated patient, while QALY gains were less than 0.025, resulting in an estimated cost per QALY of $15 000–36 000 (QI-medication) and $9500–21 500 (QI-therapy). In addition to these health gains, patients exposed to the quality improvement programmes were employed more days than those receiving usual care.

The envisaged benefit of expressing the results of economic evaluation in these terms lies in the ability to line-up cost-per-QALY estimates for a range of different interventions and disorders, with a view to determining acceptable efficiency against a pre-defined threshold (of, say, $50 000 in the US context), or even constructing 'league tables' summarizing best and worst buys in the health sector. In practice, there remain significant problems in relying on league tables for allocating resources (due to the heterogeneous and context-specific nature of cost-utility studies), while there may be criteria unrelated to efficiency that determine whether a particular intervention is deemed acceptable for reimbursement or inclusion in a defined package of basic health care.

Cost-effectiveness of interventions for child and adolescent mental health problems

There are hundreds of completed economic evaluations in the depression field, almost all confined to adults of 'working age'. But there is surprisingly little economic evidence on child and adolescent mental health interventions. A systematic review a few years ago found only 14 published economic evaluations, some of rather poor quality.[20] Common problems included small sample sizes, narrow cost measures, short follow-ups, and limited outcome measures. (Guidelines and quality checklists are available for health economist researchers and readers of their outputs.)[21]

Another drawback is that most of the completed economic studies have been undertaken in North America, the United Kingdom, and Australia. But the results of economic evaluations generally do not transfer easily from one health system to another because of differences in system structure and financing, leading to differences in relative costs. It is infeasible and certainly unnecessary to carry-out an evaluation every time a policy decision needs to be taken, but it is also difficult to assess the relevance of economic evidence from another country, especially if its mental health system is markedly different.

An example of a well-conducted cost-effectiveness analysis is the evaluation of a home-based social work intervention for children and adolescents who have deliberately poisoned themselves.[22] The researchers measured suicidal ideation, hopelessness, and family functioning as the main outcomes, and costs were based on patterns of utilization of health, education, social care, and voluntary sector services. Within a randomized controlled trial, involving 162 children aged 16 years or under, they found no significant difference in the main outcomes or costs, although parental satisfaction with treatment was significantly greater in the group that received a new social work intervention compared to those who received routine care.

In another pragmatic randomized trial, a parenting intervention for parents of children at risk of developing conduct disorder (the Incredible Years programme) was compared to wait-list controls. The perspective for cost measurement was the public sector (health, social care, special education); effectiveness was measured by reductions in intensity of behaviour problems.[23] The Incredible Years programme was more effective but also more costly. The researchers found that it would cost £1344 to bring the average child in the intervention group (in terms of behaviour intensity score) to below the clinical cut-off point. A cost-effectiveness acceptability curve was plotted to show the trade-offs between cost and effectiveness.

Sectoral cost-effectiveness analysis of mental health interventions in developing countries

Cost-effectiveness analysis can also be used to evaluate mental health programmes for whole populations (countries or even world regions). While the burden of neuropsychiatric disease is very high, the resources available to address that burden are extremely low. Given the consequent tension between the need for and the availability of mental health care, plus the fact that effective interventions do exist, the job of cost-effectiveness analysis is to show how much of the burden can be reduced or averted, by doing what, and at what cost.

Through its CHOICE project (choosing interventions that are cost-effective), WHO embarked on an initiative to assemble databases on cost-effectiveness of key health interventions in 14 epidemiological sub-regions of the world.[24] A comparative cost-effectiveness analysis of interventions for reducing the burden of major neuropsychiatric disorders formed part of this programme.[6,25,26] WHO-CHOICE advocates a 'generalized' form of cost-effectiveness analysis, in which costs and effects of current and new interventions are compared to the starting point of 'doing nothing'. Accordingly, the costs and effectiveness of pharmacological and psychosocial interventions in primary care or outpatient settings for psychiatric disorders were compared in a population model to an epidemiological situation representing the untreated natural history of these disorders. Effects are measured as disability adjusted life years (DALYs) averted (i.e. reduced burden), and costs in international dollars (I$; one international dollar should buy the same quantity of health care resources in China as in the United States).

Compared to no treatment (natural history), the most cost-effective strategy for averting the burden of psychosis and severe affective disorders in developing regions of the world is a combined intervention of first-generation antipsychotic or mood-stabilizing drugs with adjuvant psychosocial treatment delivered by community-based outpatient services, with cost-effectiveness ratio of I$4200–5500 in Sub-Saharan Africa and South Asia, rising to more than I$10 000 in middle-income regions[25] (see Fig. 7.7.1). Currently, the high acquisition price of second-generation antipsychotic drugs makes their use in developing regions questionable on efficiency grounds alone, although this situation stands to change as these drugs come off patent. By contrast, evidence indicates that the relatively modest additional cost of adjuvant psychosocial treatment reaps significant health gains, thereby making such a combined strategy for schizophrenia and bipolar disorder treatment more cost-effective than pharmacotherapy alone.

For more common mental disorders treated in primary care settings (depressive and anxiety disorders), the single most cost-effective strategy is the scaled-up use of older antidepressants (due to their lower cost but broadly similar efficacy to newer antidepressants). However, as the price margin between older and generic newer antidepressants continues to narrow, generic SSRIs should be at least as cost-effective and may therefore represent the treatment of choice in the future. Since depression is commonly recurring, there are also grounds for thinking that proactive care management, including long-term maintenance treatment with antidepressant drugs, represents a cost-effective (if more resource-intensive) way of significantly reducing the enormous burden of depression in developing regions.[27]

The purpose of such an exercise is to locate the relative position of effective and applicable interventions within a wider cost-effectiveness and priority-setting framework. Using the affordability criteria of the WHO Commission for Macroeconomics and Health,[28] this analysis indicates that (a) the most efficient interventions for common mental disorders can be considered very cost-effective (each DALY averted costs less than 1 year of average per capita income), and (b) community-based interventions for severe mental disorders using older antipsychotic and mood-stabilizer drugs meet the criterion for being cost-effective (each DALY averted costs less than three times GDP per capita). These findings therefore provide relevant information regarding the relative value of investing in neuropsychiatric treatment and prevention, and so may help to remove one of many remaining barriers to a more appropriate public health response to mental health needs.

Conclusion

Economic evaluation provides a means of comparing the costs and outcomes of mental health interventions or programmes, enabling decision makers to assess whether they offer good use of (scarce) resources. An analysis of costs alone, or indeed of outcomes alone,

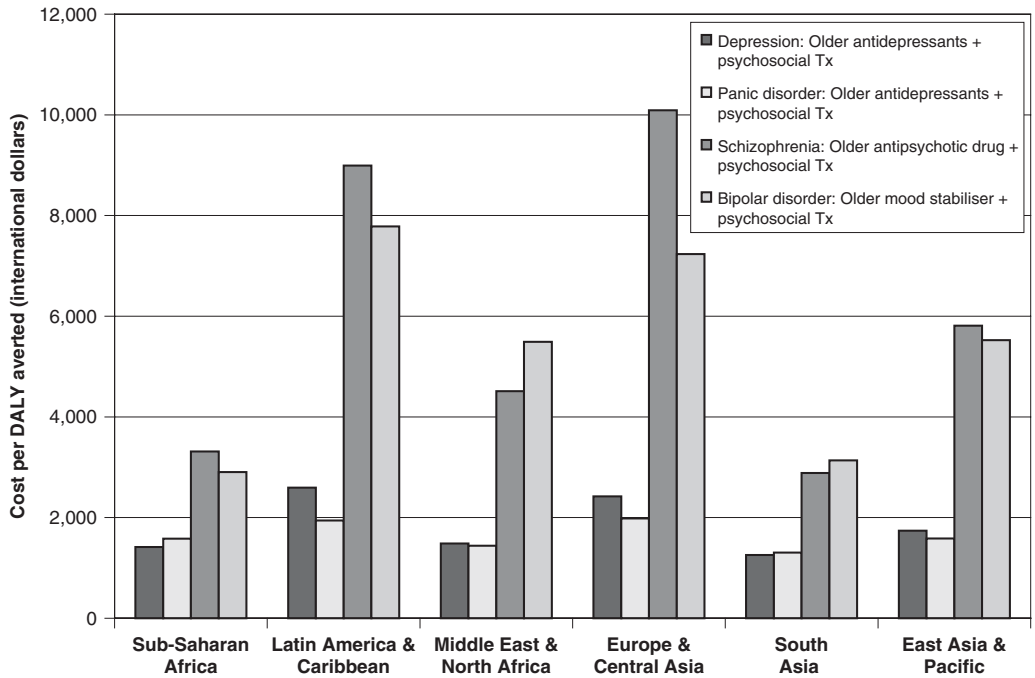


Fig. 7.7.1 Cost-effectiveness ratios for a basic mental health package in low and middle-income regions of the world.

does not provide such information. The results of well-conducted economic evaluations can be channelled into decision-making processes at a succession of levels.

Patients, users, and caregivers

Economic evidence can complement clinical decision-making at the patient, user, and caregiver level by comparing costs and consequences of particular treatments. One very pertinent question about any treatment is whether the additional acquisition costs associated with (say) newer antidepressants or second-generation antipsychotics are compensated by better symptomatic response or fewer side effects. Our earlier examples of economic evaluations of treatments for depression and for child and adolescent mental health problems provide this kind of evidence.

Purchasers and providers

At another decision-making level, those who commission or purchase mental health services need economic data. A core element of local needs assessment and strategic service development by (say) a state health care system or a health maintenance organization concerns the resource implications of changes to, for instance, the hospital/community balance or investment in a new clinic or training of therapists.

Government and society

Economic evaluations should influence national-level policy and resource allocation decisions. Such evaluations have influenced policy with respect to the substitution of community-based for long-term hospital care, the development of 'assertive outreach' models, the expansion of early intervention initiatives for psychosis, and the overall level of funding. The CHOICE programme aims to provide this kind of evidence.

While adding economic analysis to mental health evaluations introduces an extra dimension that offers a wider assessment of the implications of new or existing courses of action, there can also be limitations. Many economic evaluations fall short of the ideal, whether in terms of sample size, comprehensiveness of cost measurement, outcome assessment, or evidence interpretation. Conclusions based on small-sample randomized trials can often only be tentative, while failure to measure the wider (non-health and non-service) costs associated with two or more treatments may produce misleading and partial results.

Even when it overcomes these limitations, an economic evaluation can never resolve difficult allocative and policy issues; rather, it is one additional tool that, together with evidence on the clinical and social dimensions, can facilitate explicit evidence-based decision-making.

Further information

Knapp, M.R.J., Funk, M., Curran, C., *et al.* (2006). Economic barriers to better mental health practice and policy. *Health Policy and Planning*, **21**, 157–70. http://heapol.oxfordjournals.org/cgi/content/full/21/3/157

WHO. (2006). *Dollars, DALYs and decisions: economic aspects of the mental health system*. WHO, Geneva, Switzerland. http://www.who.int/mental_health/evidence/dollars_dalys_and_decisions.pdf

Website of WHO's cost-effectiveness work programme (CHOICE): http://www.who.int/choice

References

1. WHO. (2001). *The World health report 2001; mental health: new understanding, new hope*. WHO, Geneva.
2. WHO. (2004). *Mental health policy, plans and programmes. Mental health policy and service guidance package*. WHO, Geneva.
3. WHO. (2000). *The world health report 2000; health systems; improving performance*. World Health Organization, Geneva.
4. Dixon, A., McDaid, D., Knapp, M., *et al.* (2006). Financing mental health services in low- and middle-income countries. *Health Policy and Planning*, **21,** 171–82.

5. Knapp, M.R.J., Funk, M., Curran, C., *et al.* (2006). Economic barriers to better mental health practice and policy. *Health Policy and Planning*, **21**, 157–70.
6. WHO. (2006). *Dollars, DALYs and decisions: economic aspects of the mental health system*. WHO, Geneva.
7. Scott, S., Knapp, M., Henderson, J., *et al.* (2001). Financial cost of social exclusion: follow-up study of antisocial children into adulthood. *British Medical Journal*, **323**, 191–4.
8. Beecham, J.K.J. and Knapp, M.R.J. (2000). Costing psychiatric interventions. In *Measuring mental health needs* (2nd edn) (eds. G. Thornicroft, C. Brewin, and J.K. Wing), pp. 200–24. Gaskell, London.
9. Curtis, J. and Netten, A. (2005). *Unit costs of health and social care*. Personal Social Services Research Unit, University of Kent, Canterbury.
10. McCrone, P., Knapp, M., Proudfoot, J., *et al.* (2004). Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. *The British Journal of Psychiatry*, **185**, 55–62.
11. Chisholm, D., Healey, A., and Knapp, M.R.J. (1997). QALYs and mental health care. *Social Psychiatry and Psychiatric Epidemiology*, **32**, 68–75.
12. Olsen, J.A. and Smith, R.D. (2001). Theory versus practice: a review of 'willingness-to-pay in health and health care. *Health Economics*, **10**, 39–52.
13. Valenstein, M., Vijan, S., Zeber, J.E., *et al.* (2001). The cost-utility of screening for depression in primary care. *Annals of Internal Medicine*, **134**, 345–60.
14. Revicki, D., Brown, R., Keller, M., *et al.* (1997). Cost-effectiveness of newer antidepressants compared with tricyclic antidepressants in managed care settings. *The Journal of Clinical Psychiatry*, **58**, 47–58.
15. Kamlet, M.S., Wade, M., Kupfer, D.J., *et al.* (1992). Cost-utility analysis of maintenance treatment for recurrent depression: a theoretical framework and numerical illustration. In *Economics and mental health* (eds. R.G. Frank and W.G. Manning), pp. 267–91. Johns Hopkins University Press, Baltimore, MD.
16. Hatziandreu, E.J., Brown, R.E., Revicki, D.A., *et al.* (1994). Cost-utility of maintenance treatment of recurrent depression with sertraline versus episodic treatment with dothiepin. *Pharmacoeconomics*, **5**, 246–64.
17. Pyne, J.M., Smith, J., Fortney, J., *et al.* (2003). Cost-effectiveness of a primary care intervention for depressed females. *Journal of Affective Disorders*, **74**, 23–32.
18. Schoenbaum, M., Unutzer, J., Sherbourne, C., *et al.* (2001). Cost-effectiveness of practice-initiated quality improvement for depression; results of a randomized clinical trial. *The Journal of the American Medical Association*, **286**, 1325–30.
19. Knapp, M., Romeo, R., Mogg, A., *et al.* (in press). Cost-effectiveness of transcranial magnetic stimulation vs. electroconvulsive therapy for severe depression: a multi-centre randomised controlled trial. *Journal of Affective Disorders*.
20. Romeo, R., Byford, S., and Knapp, M. (2005). Annotation: economic evaluations of child and adolescent mental health interventions: a systematic review. *Journal of Child Psychology and Psychiatry*, **46**, 919–30.
21. Drummond, M.F. and Jefferson, T.O. (1996). Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ economic evaluation working party. *British Medical Journal*, **313**, 275–83.
22. Byford, S., Harrington, R., Torgerson, D., *et al.* (1999). Cost-effectiveness analysis of a home-based social work intervention for children and adolescents who have deliberately poisoned themselves. Results of a randomised controlled trial. *The British Journal of Psychiatry*, **174**, 56–62.
23. Edwards, R., Ceilleachair, A., Bywater, T., *et al.* (2007). Parenting programme for parents of children at risk of developing conduct disorder: cost effectiveness analysis. *British Medical Journal*, **334**, 682–7.
24. Tan Torres, T., Baltussen, R.M., Adam, T., *et al.* (2003). *Making choices in health: WHO guide to cost-effectiveness analysis*. WHO, Geneva.
25. Chisholm, D. (2005). Choosing cost-effective interventions in psychiatry. *World Psychiatry*, **4**, 37–44.
26. Hyman, S., Chisholm, D., Kessler, R., *et al.* (2006). Mental disorders. In *Disease control priorities in developing countries* (2nd edn) (eds. D. Jamison, J. Breman, A. Measham, *et al.*). Oxford University Press, New York.
27. Chisholm, D., Sanderson, K., Ayuso-Mateos, J.L., *et al.* (2004). Reducing the global burden of depression: a population-level analysis of intervention cost-effectiveness in 14 epidemiologically-defined sub-regions (WHO-CHOICE). *The British Journal of Psychiatry*, **184**, 393–403.
28. Commission on Macroeconomics and Health. (2001). *Macroeconomics and health: investing in health for economic development*. WHO, Geneva.

7.8

Psychiatry in primary care

David Goldberg, André Tylee, and Paul Walters

Epidemiology

In recent years, major epidemiological surveys have been carried out in the community in many different countries, in the United Kingdom most recently by the Office of National Statistics (to find this and other surveys go to http://www.statistics.gov.uk/STATBASE/Product.asp?vlnk=8258). The findings in such community surveys can be compared with findings in primary care surveys, when it will be found that the list of common mental disorders is not quite the same, although conditions characterized by symptoms of depression and anxiety are the most common. Rather than considering the detailed diagnoses, it can be helpful to distinguish between 'internalizing disorders', which besides anxiety states and depression, also include the fear disorders like phobias and panic disorder, obsessive-compulsive disorder and many cases of somatization disorder; and 'externalizing disorders' consisting of conduct disorder in childhood, and antisocial behaviour, as well as drug and alcohol disorders in adult life. The former group of disorders are characterized by subjective distress, and typically high levels of anxious and depressive symptoms; while in the latter group abnormalities are in externally observed behaviour.[1]

In community surveys it can be seen that rates of internalizing disorders rise sharply after puberty, are highest between the ages of 35 and 55 and fall thereafter, and that females rates are higher than males at all ages., while rates of externalizing disorders reach their maximum between the ages of 15 and 34, and fall sharply after that, with males rates much higher than female rates at all ages. This is shown in Table 7.8.1 (which does not include antisocial behaviour as reliable data are not available in the community).

The Goldberg-Huxley Model[2]

This was devised as a framework for comparing the characteristics of patients seen in the community with those in other medical settings, and describing the pathway which people usually follow to mental health care in places where GPs act as 'gatekeepers'. It consists of five levels, separated by four filters. The figures for psychiatric morbidity over 1 year necessitate using estimates of incidence rates, and are therefore much higher than the point prevalence rates reported in community surveys. The essence of the model is the demonstration that most distressed patients will see a doctor over the course of 1 year (filter 1), but only about half of them will have their distress detected (filter 2). Most common mental disorders are treated in primary care, so filter 3 is relatively impermeable, only allowing one in five to pass. Psychiatrists only have any part in the process with the fourth filter, which also holds back most patients. Psychiatrists therefore form their ideas about mental disorders from a highly skewed section of all those with disorders.

Prevalence of psychiatric disorder in primary care

In the United Kingdom, about 80 per cent of the population consult their doctor in the course of a year, and prevalences among attenders are higher than among the general population.[2] In contrast, specialist mental health services see between 1 and 2 per cent of the population in the course of a year, and admit only about 0.5 per cent to inpatient care, so that primary care deals with the major part of the burden of common mental disorders.

The World Health Organization (WHO) carried out the largest primary care survey in 14 countries[3] but for purposes of comparison only the UK data will be shown here. Table 7.8.2 compares the frequencies and types of mental disorders seen in the community, in primary care, and in psychiatric practice. Mental disorders

Table 7.8.1 Annual prevalence of mental disorders in the community by type and age, rates per 100 at risk

Disorder	Gender	5 to 16	to 34	to 54	to 74	All (16–75)
Internalizing	Male	3.1	11.70	16.75	9.87	13.5
	Female	4.3	20.55	21.35	14.80	19.4
Externalizing	Male	10.05	11.6; 18.9	2.25; 10.4	0.4; 3.8	6.0; 11.9
	Female	4.35	5.3; 5.7	0.75; 2.1	0.4; 0.5	2.3; 2.9
Other	Male	1.9	0.33	0.78	0.23	0.5
	Female	0.75	0.42	0.73	0.52	0.6

Internalizing = any neurotic disorder. *Externalizing* = conduct disorder for age 5–16; for the remaining age groups the rate for drug dependence is shown first, followed, after the semicolon, by the rate for alcohol dependence. *Other* = psychotic disorders in adults. Source: National Statistics website: www.statistics.gov.uk Crown copyright material is reproduced with the permission of the Controller Office of Public Sector Information (OPSI).

Table 7.8.2 Prevalence of mental disorder by gender for the community, for primary care attenders, and for admissions to psychiatric beds

	The community annual prevalence (%)		Primary care cases consecutive attenders (%)		Mental hospital inpatients (%)	
	Males	Females	Males	Females	Males	Females
Mixed anxiety depression	6.8	10.8	2.1	4.5	9.8	17.6
GAD	4.3	4.6	4.9	14.9		
Panic	0.7	0.7	3.4	3.6		
Phobias	1.3	2.2	2.1	4.6		
Neurasthenia	–	–	6.1	21.7		
Somatoform disorder	–	–	–	0.5		
OCD	0.9	1.3	–	–		
Depression	2.3	2.3	13.9	18.3	17.9	27.3
Alcohol dependence	11.9	2.9	5.3	0.8		
Drugs dependence	5.4	2.1	–	–	30.1	14.3
Schizophrenia	0.6	0.5	–	–	20.4	13.7
Organic, dementia	–	–	–	–	10.3	15.9
Subnormality	–	–	–	–	7.1	5.4
Developmental disorders	–	–	–	–	5.1	5.4
Any Dx	14.1%	19.9%	23.5%	27.5%	100%	100%

Sources: National Statistics website: www.statistics.gov.uk Crown copyright material is reproduced with the permission of the Controller Office of Public Sector Information (OPSI).

seen in primary care settings are more severe on average than those seen in community surveys, and different disorders predominate. The figures shown are for practices in Manchester with a fairly high prevalence of mental disorders, but the spread of diagnoses is fairly similar in other countries. The ICD-10 criteria only counted somatoform disorders if they were severe and long-standing, and do not count the many patients presenting with unexplained somatic symptoms, which are often accompanied by symptoms typical of anxiety or depression. A more recent study from Denmark[4] has estimated that almost a half of their patients were diagnosed cases of mental disorders, with somatoform disorders being found in about one-third. Patients with established physical illnesses are also at greater risk of mental disorders, and this is especially so if they are disabled by their illness. It can be seen from Table 7.8.3 that disorders admitted to psychiatric hospitals in the United Kingdom are different again from those typically seen in primary care, with organic states, drug and alcohol dependence, schizophrenia and severe depressive states accounting for the majority of cases (Source: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=202).

A study in 10 European countries shows 28 per cent of consecutive attenders in the United Kingdom to be distressed on a screening interview, but only 6 per cent presented psychological symptoms to their GP. Most of these (5.5 per cent) received a psychiatric diagnosis, but the GPs also diagnosed others as 'psychiatric'—so that their total rate was 15 per cent.[5] These figures are fairly similar to those in Switzerland and the Netherlands, but in stark contrast to those in Eastern Europe. In the Russian Federation, for example, 27 per cent were distressed, but none reported psychological distress to their GPs, and none were diagnosed: however, the GPs identified 3 per cent of their patients as 'psychiatric'. Fairly similar figures were reported in Estonia, Poland, and Belgium; while figures in Germany, Spain, and Sweden are intermediate (*ibid* 2007; see Fig. 7.8.1). This study also showed that GPs who discuss psychosocial matters with their patients, and look at them are better at diagnosing them—a finding that echoes previous research in the United Kingdom.[2]

It can be seen from Fig. 7.8.1 that most distressed patients—who may well be found to have a mental disorder if interviewed with a research interview—do not mention their distress to their doctor, and that this accounts for failure to diagnose the disorder. Many patients who have not endorsed feelings of distress are nonetheless assessed as mental unwell, either because they are presenting with

Table 7.8.3 The Goldberg-Huxley Model: Data for Manchester, UK

Level 1: Community samples *First Filter: The decision to consult*	**250–315/1000/year**
Level 2: All those seeing GPs found to have a mental disorder *Second filter: GPs ability to detect*	**210–230/1000/year**
Level 3: Cases recognized by the GP As 'mental disorders' *Third filter: GPs decision to refer*	**101/1000/year**
Level 4: All those seeing mental health professionals *Fourth filter: Psychiatrist's decision to admit to hospital*	**20.6/1000/year**
Level 5: Those admitted as inpatients	**3.4/1000/year**

(Note that these estimates are annual period prevalences, and depend on estimates of annual incidence rates in addition to point prevalence rates. Source: Reproduced from D.P. Goldberg and P.J. Huxley, *Common mental disorders—a bio-social model*, copyright 1992, The Tavistock Institute, London.)

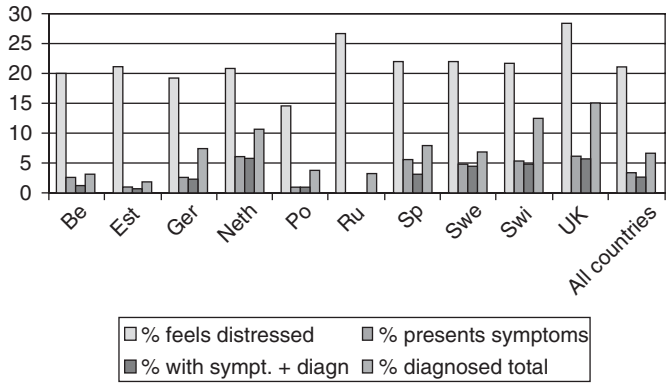


Fig. 7.8.1 Proportions of population that feel distressed, present symptoms of distress, are diagnosed if they do so, and total number of patients diagnosed, in 10 European countries (Reproduced from P. Verhaak, J. Bensing, and A. Brink-Nuinen, Primary mental health care in 10 European countries: patients' demands and GPs' responses. *European Journal of Psychiatry*, **21**, (1), Zaragoza Jan–Mar 2007, copyright 2007, INO REPRODUCCIONES, S.A.).

unexplained somatic symptoms or because they are under treatment for a mental disorder that has responded to treatment.

Clinical presentations

Somatization

Somatization is broadly defined as the expression of psychological distress through physical symptoms. In primary care, most patients present physical symptoms at the onset of an episode of anxiety and depression. This is actually the usual way that new episodes of common mental disorders present in general medical settings, as only about 15 per cent of new episodes present purely in psychological terms.[2]

Primary care somatizers can be subdivided into 'facultative somatizers', who admit to their psychological symptoms and accept a mental disorder diagnosis if appropriately interviewed, and 'pure somatizers' who, despite such an enquiry, still deny the presence of psychiatric symptoms. People have many reasons for preferring to present somatic symptoms to their doctors—they are understandably worried that they may have a new physical disease, they give priority to pains over other symptoms because it is pain that hurts, and they often wish to avoid the stigma of being thought psychologically distressed.

Somatizers may be compared with 'psychologisers', who directly present their psychological problems to their doctors. The former are more likely to report adverse childhood events and periods of childhood illnesses, while the latter have more abnormal attachment behaviours.

Hidden versus conspicuous morbidity

The ability of GPs to detect psychological disorders among their patients forms the second filter that patients must pass through in order to receive a diagnostic label. This ability varies a good deal between places as well as between disorders. In the WHO study[3] the overall average detection rate was 48.9 per cent, but rates varied from 75 per cent for Verona to 15.9 per cent for Shanghai (62.9 per cent for Manchester). Recognition rates for individual diagnoses followed these overall rates, with somatization disorder being best recognized, followed by depression. However, these detection rates do not reveal whether the GP is identifying the same patients as the research assessment. In fact, the exact agreement between the two (measured by κ) is rather poor, at only +0.18 for all centres (+0.38 for Manchester).

The second filter is passed when the GP recognizes a mental health problem in the patient, although this will often be without a precise ICD-10 diagnosis. Those recognized by the GP make up the '**conspicuous' morbidity**—in fact, just under half of that estimated to be present in the waiting room population by a two-stage case-finding procedure: so that the patients who are not identified can be thought of as the '**hidden morbidity'**. These undetected patients continue to consult, but an outsider's inspection of notes and prescriptions, or even discussion with the relevant doctor, will not identify them as patients with psychiatric morbidity. In practice, the 'conspicuous' morbidity may be greater than 50 per cent, as the longitudinal nature of primary care means that patients may be diagnosed in subsequent visits, and this is missed in cross-sectional waiting room studies.[6,7] Kessler *et al.* followed up a cohort of primary care patients over 3 years and found only 14 per cent of patients with depression remained unrecognized at the end of this period.[8] Rost *et al.* followed up 98 depressed patients who had made at least one visit to their GP and found 32 per cent were undetected at 1 year.[9] Despite this, GPs are good at recognizing severe depression, and unrecognized depression tends to be mild.[10–12] The severity of depression in primary care, rather than being defined categorically, may therefore be better conceptualized as running along a continuum from mild to severe. Using a dimensional approach Thompson *et al.* calculated GPs only miss one 'probable' case of depression every 29 consultations.[13]

Doctors better able to detect disorder have the following characteristics:[2]

- Make eye contact with the patient
- Make empathic comments
- Pick up verbal cues
- Pick up non-verbal cues
- Ask directive questions, with a psychological content
- Do not read notes, or look at their computer, while the patient is speaking
- Deal with over-talkativeness
- Deal with today's problem

Data from the WHO study indicate that these 'undetected illnesses' are on an average less severe than those detected by GPs and have a somewhat better outlook. However, the data does not support the view that failure to detect these less severe disorders has serious long-term consequences for the patient[14] although this does not mean that there are not individuals would be better served if their distress was acknowledged.

The elderly

Mental illness in the elderly is common in primary care. Between 5 and 10 per cent of older adults attending primary care will suffer from depression, though this may be higher in areas of socio-economic

deprivation.[15] Older people are less likely to admit to psychiatric problems and more likely to emphasize somatic concerns and present behavioural changes. One study found that only 38 per cent of those identified as depressed through screening in the community had discussed feelings of depression with their GP.[16] This may lead to under-recognition in primary care. Crawford *et al*. found that only 52 per cent of 62 patients identified with clinical depression from a community survey were correctly diagnosed by their GP.[17] The elderly may attribute their symptoms to 'normal ageing', grief, or physical illness, or may fear sigmatization more than younger patients making recognition more difficult. Elderly depressed men may be particularly likely to go unrecognized.[17]

A systematic approach using a collaborative care model may improve depression management for the elderly in primary care. The Improving Mood-Promoting Access to Collaborative Treatment (IMPACT) trial in the United States demonstrated that a primary care collaborative model for late-life depression was more effective than usual care in improving depressive symptoms. It also decreased pain due to osteoarthritis, increased functional abilities, and improved quality of life.[18] The collaborative care model was highly cost-effective[19] and continued to show benefits over a 2-year follow-up period.[20]

The other major condition with which the GP will be involved is dementia. In the United Kingdom, the prevalence rate is approximately 5 per cent for all those over 65, but there is an age-related rise within this band to 25 per cent for those aged over 85. Consultations for organic psychoses reflect this: 370 consultations per 10 000 years at risk for those aged over 75, and 888 per 10 000 years at risk for those aged over 85. However, the management of people with dementia in primary care has been criticized.[21] Up to 75 per cent of patients with moderate to severe dementia and up to 97 per cent of patients with mild cognitive impairment go unrecognized by their GP.[22] Again there may be a number of reasons for this. Dementias have an insidious onset, and sometimes the doctor's familiarity with the patient can militate against spotting change. If a relative also accepts that the changes associated with dementia reflect normal ageing, the diagnosis may be delayed or never made. Only 40 per cent of GPs in the United Kingdom use a specific test to detect dementia, and in a survey of 8051 GPs in England 40 per cent thought an early diagnosis of dementia was not important.[23] Turner *et al*. surveyed GPs knowledge, confidence, and attitudes about dementia and found that despite GPs overall knowledge about diagnosis and management being good, a third lacked confidence in their diagnostic skills and two-thirds lacked confidence in their management of behavioural problems and other associated problems in dementia.[24]

Downs and others conducted a trial of an educational package to improve detection and management of dementia in primary care.[25] In the United States, a trial of collaborative care versus care as usual for patients with dementia in primary care has produced encouraging results.[26] Compared with care as usual, collaborative care (consisting of case management though a senior practice nurse working with the patient's family and integrated in the primary care team, and the use of standard protocols to guide and monitor treatment) resulted in significant improvements in the quality of care and in the symptoms of dementia, without an increase in psychotropic medication use.

Classification of mental disorders in primary care

Difficulties with conventional psychiatric taxonomies

The main problem with the International Classification of Diseases, 10th Edition (ICD-10)[27] or Diagnostic and Statistical Manual, 4th Edition (DSM-IV)[28] classifications used by psychiatrists is that they were devised to describe a very different consulting population, they are needlessly complicated, and they do not lead directly to management. Patients usually present a mixture of physical, psychological, and social symptoms expressed in any order, although somatic symptoms are usually first. Some symptoms are repeatedly mentioned and some are mentioned only in passing. Symptoms left to the end may be the most important of all. Symptoms may not fit a psychiatrist's taxonomy. In primary care, patients often have several concurrent problems of a medical, psychological, and social nature.

Most psychologically distressed patients show symptoms of both anxiety and depression. The ICD-10 classification has a mild disorder called 'mixed anxiety depression', since some patients have symptoms of each which together seem sufficient for a diagnosis, although not satisfying the criterion for either disorder on its own. However, this does not solve the problem of the many patients who are above the threshold for both disorders, who are declared 'co-morbid' for two different disorders by conventional psychiatric taxonomy.

An alternative view points out that the two groups of symptoms are strongly correlated with one another (about +0.7) in the consulting population,[2] and thus views them as two related dimensions of symptomatology which tend to co-vary over time. GPs themselves rarely emphasize the distinction between the two groups of symptoms, and there appears to be no evidence that any adverse consequences follow this neglect. For the GP, the diagnostic task can be one of separating the symptoms of depression and anxiety from those of an accompanying physical illness, or of probing for psychiatric morbidity in patients where apparent physical symptoms do not have an organic cause. In primary care settings, 'co-morbidity' refers to patients who have both physical disorders and mental disorders.

Solutions to the classification problem

Both the major psychiatric classifications—ICD-10 and DSM-IV—offer special versions produced in collaboration with primary care physicians which are deemed suitable for this setting, and roughly correspond to the parent classification. The WHO offers '**ICD10-PHC**'[29] which consists of 26 common conditions, with advice on how they present, the diagnostic features, the differential diagnosis of similar symptoms, essential information for the patient and the carer, advice and support for the patient and the carer, the role of medication, and indications for referral to the mental health services. Information is given about national organizations, and self-help materials, to assist people with particular diagnoses, and advice is given to GPs in particular areas about customizing the system by including information about self-help and support groups for local people.

The '**DSM-IVPC**'[30] on the other hand is organized by symptoms that branch out into diagnostic algorithms. The GP assesses the

patient's symptoms and, in workbook fashion, determines the relevant psychiatric diagnoses. The manual is formatted to be concise and practical, with limited use of psychiatric jargon. As a key feature, the chapter devoted to 'Algorithms for Common Primary Care Presentations' presents nine algorithms, headed by the presenting symptoms, for the most common psychiatric concerns encountered in primary care. No advice is given on management or on information to be given to carers.

Given the quantity of printed material that floods into GPs offices, it is doubtful whether many GPs keep such systems on their desks—although they may be incorporated in their computer programs. It is more realistic to use such systems in training new doctors, so that ways of approaching patients and their carers become incorporated in their usual routines, with the system only consulted when an unusual problem presents itself. It can be seen that whereas the DSM-IV system is aimed at formal diagnosis, the ICD10-PHC system is aimed at management once an assessment has been completed.

The International Classification for Primary Care, (ICPC-2-R)[31]

The classification most widely used by GPs is of course their own, devised under the auspices of the World Organization of National Colleges & Academies ('WONCA'), called the International Classification for Primary Care, '**ICPC-2-R**'. This is a system which classifies all patient data and clinical activity in primary care, taking into account the frequency distribution of problems commonly encountered.

It allows classification of

- the patient's reason for encounter (RFE),
- the problems/diagnosis managed,
- interventions,
- test results, and the
- ordering of these data in an episode of care structure

It has a biaxial structure and consists of 17 chapters, each divided into seven components dealing with

- symptoms and complaints
- diagnostic, screening, and preventive procedures
- medication, treatment, and procedures
- test results
- administrative
- referrals and other reasons for encounter, and
- diseases

It is not clear to what extent all GPs work their way through this complex system, but should they reach the seventh component, there is a rough correspondence with the ICD. Note that multiple disorders can be coded in the same episode, but the extent to which this is done is not clear. Nor does the system provide advice on the management of the various conditions—it is assumed that the clinician knows how to do this.

The Read codes[32]

The most widely used system in the United Kingdom since the advent of computerization, are the Read codes. This meets the needs of the generalist by including diagnoses, symptoms, and problems. Some very broad reasons for consultation—such as 'anxiousness', 'depressed', and 'headache' are provided. A letter of the alphabet is followed by up to four numerical codes, and there are also about 50 codes for diagnoses such as E204. 'Depression' or E2003 'anxiety with depression', which correspond very roughly to ICD diagnoses. However, no criteria are given, nor any advice on management for these various diagnoses.

None of the above four systems take into account functional impairment and disability, yet clinicians need to consider this in conjunction with the set of symptoms presented by an individual patient. Both diagnosis and current impairment are essential, and may help to explain why up to a quarter of patients with schizophrenia are managed solely in primary care settings[33] whilst some patients with adjustment disorder need referral to the community mental health team.

Improving the identification of some common disorders

(a) Aids to accurate detection of depression

Rather than using routine screening questionnaires to all patients, it is more practical to ask the following questions in six groups of patients:

- all those who look or sound depressed, or mention depressive symptoms
- all those with a past history of depression
- all those with significant physical illness causing disability
- all those with diabetes and coronary heart disease, where the risk is higher
- all those with other mental health problems, such as dementia or heavy drinking
- mothers of infants who are either single or are unsupported by a partner or family

Both these questions should be asked:

- During the past month, have you been feeling down, depressed or hopeless?
- During the last month, have you often been bothered by having little interest or pleasure in doing things?

(b) Aids to accurate assessment of severity of depression

If positive replies are obtained to either of these, assess severity of using the Patient Health Questionnaire-9 (PHQ-9) a questionnaire with nine questions validated for use in primary care,[34] which may assist in ensuring that antidepressant medication is better targeted on those with moderate and severe degrees of depression.

(c) Aids to the accurate detection of alcohol problems

Other, more recent tools, such as the Drug Abuse Problem Assessment for Primary Care (DAPA-PC),[35] were designed specifically to be administered via computer. While currently popular in the areas of depression and substance abuse, the audience for assessment tools is expanding for a number of reasons, including: convenience, privacy, high-patient satisfaction,[36] decreased provider time, improved validity, and reliability,[37] and decreased expense.

The management of mental disorders within primary care

(a) Stepped care

With the introduction of mental health guidelines there has been a shift towards stepped care for mental health problems.[38] Stepped care 'provides a framework in which to organize the provision of services supporting both patients and carers, and healthcare professionals in identifying and accessing the most effective interventions' (see Fig. 7.8.2). Stepped care allows treatment to be provided in steps according to the severity of problems and/or response to treatment, aiming to provide the greatest benefit to most of the people from the resources available. Patients who fail to respond are given the next step on the treatment plan. Stepped care can be delivered by starting at the least invasive step and gradually 'climbing' the steps according to response, thus targeting more intensive treatments to those that that need them, or by 'stratifying' care so that the first treatment step is determined by severity of disorder and thereafter by response.[39] Stepped care was developed in the United States, initially for the management of depression.[40] In the United Kingdom, this model of care has now been recommended for depression,[38] anxiety disorders,[41] obsessive-compulsive disorder,[42] and self-harm.[43]

(b) Mental health workers based in primary care

Several new possibilities have emerged recently with the advent of the new *graduate mental health workers* in primary care. The NHS plan (DH 2000) was to have 1000 such workers in England and indications are that there are 600–700 in post. Mostly, the posts are similar to assistant psychologists and they generally are trained on a one-day release scheme at local university level in primary care mental health. Many are conducting initial assessment, providing brief psychological interventions such as brief cognitive behaviour therapy (CBT) or problem-solving for common mental health problems and medication management. Many are overseeing the use of self-help written materials and computerized help such as 'Beating the Blues', a computerized CBT programme for depression, which has been shown to be cost-effective.[44] Graduate primary care mental health workers (PCMHWs) also act as advisors on the range of other local services (for example, local library bibliotherapy schemes, and support groups). They have been proved to be effective at increasing patient satisfaction with episodes of care but were found not to improve mental health symptoms or to use the voluntary sector more than usual care.[45] Many PCMHWs are psychology graduates, and they tend to move on after a year or two into much wanted clinical psychology posts, so corporate memory can be diminished. Many other interventions described in level two of the National Institute for Clinical Excellence (NICE) Depression guideline[38] are increasingly applied in primary care—such as exercise schemes, befriending schemes, brief problem-solving, brief CBT, self-help materials, and sleep restoration.

Practice primary care counsellors may in reality be psychologists, psychotherapists, or counsellors depending on their training, expertise, and accreditation. They mostly see adults with common mental health disorders which may include adjustment disorders and losses. They often integrate different models of brief psychotherapeutic treatment depending on their training and the particular patient, and they receive regular supervision. These services may be provided in-house or in a neighbouring practice depending on

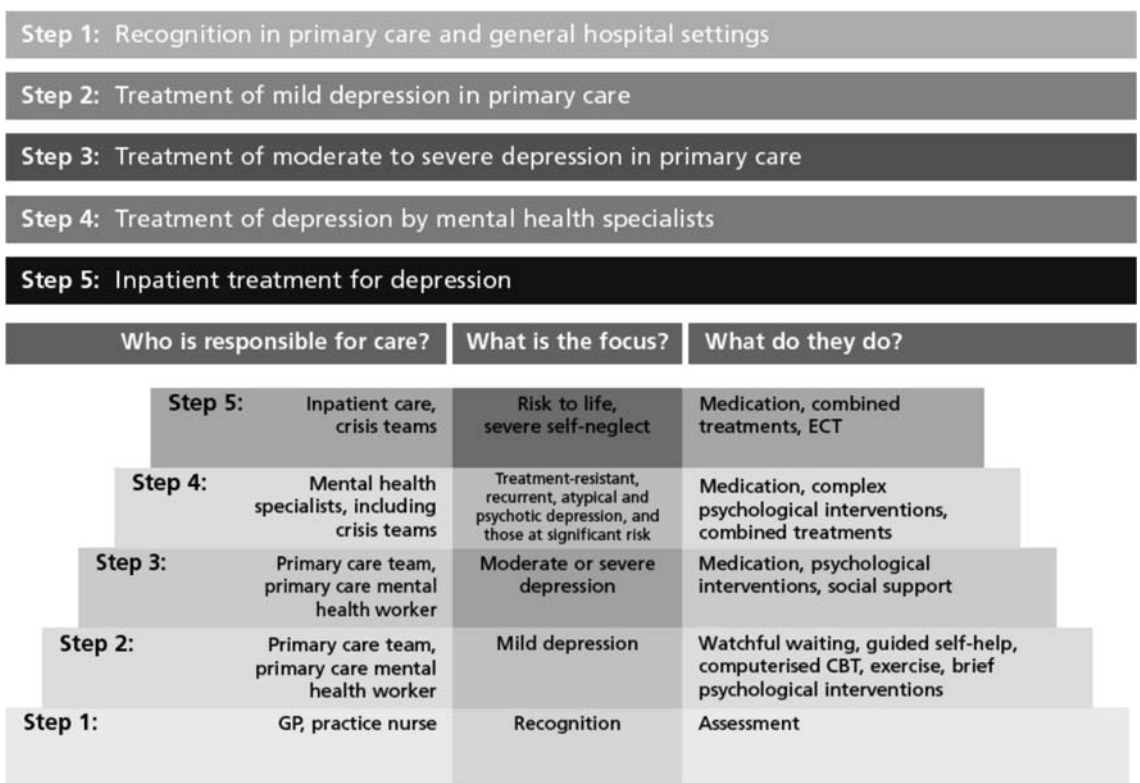


Fig. 7.8.2 The stepped care model applied to depression. Reproduced with permission.[38]

available space. Some counsellors specialize in providing services for non-English speaking groups and certain patient subgroups such as eating disorders, domestic violence, etc.). Whilst resources are usually limited on occasions they may see patients for longer courses of treatment.

(c) The role of the primary care nurse

Practice nurses are increasingly involved in chronic disease management within practices and this includes managing mental health conditions whether severe or common. From April 2006, practices have been rewarded in England under the Quality and Outcomes Framework of the General Services Medical Contract[46] for screening their patients with coronary heart disease (CHD) and diabetes for depression and with any newly depressed patients for using a recommended patient self-report instrument to assess baseline severity of depression. The recommended instruments are the Patient Health Questionnaire-9 item (PHQ9),[34] the Hospital Anxiety and Depression Scale (HADS),[47] or the Beck Depression Inventory (BDI).[48] Practice nurses are expected to undertake much of this work as they see the patients with CHD or diabetes for review and therefore they need proper training and supervision. Also rewarded in the GP contract is an annual physical review of all patients on the Severe and Enduring Mental Illness registers (SEMI or SMI)—again usually conducted by practice nurses. Such nurses have traditionally administered depot medication for such patients without much training on the procedure or the accompanying health checks. This training has been improving lately. Expecting an untrained practice nurse to administer depot without support and training is unacceptable, and training is needed in how to give the injection, how to review the suitability of this form of medication, how to monitor for side-effects, how to assess mental state for signs of deterioration, when to involve colleagues or refer to the community mental health team, and crisis management.

Practice nurses have been successfully trained in the assessment and management of depression[49,50] and the use of problem-solving in major depression.[51] Practice nurses have been successfully trained in chronic disease management of the severely mentally ill,[51] whereby they constructed a disease register, a care plan, and 3-monthly structured reviews encompassing psychological, physical, and social well-being. Whilst the practice nurses were diligent and detected problems, there was little evidence that they communicated these to the GPs. There was also a failure to increase health promotion in this group, despite their recognized increased standardized mortality.

Health visitors trained in counselling have been shown to benefit women with postnatal depression,[52] and Appleby *et al.*[53] have targeted the same group of patients with a nurse-led intervention.

Primary care mental health services

The primary care team provides the majority of care to people suffering from mental health problems. Though many services remain GP-led, many are becoming increasingly sophisticated with a growing number of professionals involved in the care of people with mental health problems. In England, the Department of Health has produced a guide for improving primary care mental health services.[39]

A number of models for primary care mental health services are described reflecting the need for services to provide care to populations with different needs, resource availability, and relationships with secondary care. Examples include the 'opt-in' or self-referral model, proactive models such as a collaborative care model, 'one-stop shop' models, and focused service models (i.e. services are developed for a specific remit such as common mental disorder or psychosis). The 'opt-in' model is one in which the patient initiates care and follow-up and is the traditional primary care model. In a collaborative care model there is active follow-up and collaboration between patient, primary care health professionals, and when required secondary care professional with the health professional coordinating care. A 'one-stop shop' model provides a single point of access to all mental health services with a triage system determining resource allocation. There is no perfect model, and all have advantages and disadvantages depending on the population needs and availability of resources.

As services develop so the range of primary care mental health professionals is expanding. As well as the traditional primary care professionals—the GP and practice nurse—many practices now have others, including primary care graduate mental health workers and health visitors. Their roles include making assessments, facilitating referrals and movement through the health care system, and delivering treatments. In the United Kingdom, the availability of counselling services is now commonplace in primary care and up to 50 per cent of general practices have on-site counselors.[54] They are referred people with a wide range of common mental disorders and other psychosocial problems such as adjustment disorders. Bower and Roland reviewed the evidence for the effectiveness of counselling in primary care and found it was associated with a modest improvement in short-term outcomes compared to usual care, but was no better than usual care over the longer-term. It did not appear to be any more cost-effective than usual care but patients appeared satisfied with it.[55]

The primary–secondary care interface

What GPs expect of psychiatrists

GPs expect psychiatrists to possess and exhibit specialized skills of assessment and management not possessed within the primary health care team. The primary care team may include primary care mental health workers ranging from graduate workers to counsellors, psychotherapists, or psychologists, although if present these are usually extremely thinly spread (i.e. 1–2 sessions per practice per week). GPs do expect their psychiatrists or a named person from the mental health team to be available when they are needed, and they should also make themselves available for incoming telephone calls from the key contact. Because of sheer numbers (in England and Wales there are 12 times as many GPs as psychiatrists) the GP must protect this valuable resource by not overloading it with inappropriate referrals and by obtaining and maintaining certain assessment and management skills that can be used in primary care as well as sharing the care of certain patients under the leadership of secondary care. Referrals will vary widely from practices depending on their own 'in-house' expertise and the presence or absence of any primary care mental health workers or local counselling services. Larger practices often have more in-house expertise and are more likely to have at least one of their GPs who possess more specialized mental health skills. Small practices with one or two partners are less likely to have such skills and may need

more support. GPs expect the psychiatrist to provide inpatient care when needed (e.g. serious self-neglect, suicide intent, etc.) and day-patient facilities to provide a place of care, respite, and safety. They can also expect the psychiatrist to use diagnostic facilities and investigations (e.g. scans) as necessary and to provide highly specialized treatments when indicated (e.g. electroconvulsive therapy). Less frequently, the GP may need respite from particular doctor–patient relationships for the longer-term good of both. Referral is also sometimes a result of pressure by the relatives or patient. Where good communication exists between primary and secondary care, these often 'covert' reasons for referral can be openly discussed.

Access to specialist assessment when appropriate is paramount for a primary care service. GPs are not usually trained in specialist assessment and therefore, to match need to services, a psychiatrist, community psychiatric nurse, or psychologist from the community mental health team can perform this function, often in a primary care setting or the patient's home. Other community mental health teams operate an outpatient clinic (which may be moved into the surgery). Often 'true consultancy' is being sought by the GP, whereby he or she may receive advice only. Other practices operate a joint consultation system whereby the specialist and generalist see the patient together and formulate a plan. With recent changes in the configuration of mental health teams in the United Kingdom to provide acute care, early intervention and assertive outreach, in some places the more traditional community mental health teams are being reduced which is often confusing and can be unsettling for practices which have developed relationships with individual psychiatrists and colleagues. Also, if crisis teams work 9 to 5 p.m., yet the GP surgeries are open until 6 or 7 p.m., this can create difficulties for making urgent referrals.

Ways of organizing the interface

The interface between primary and secondary care is of key importance in the delivery of mental health care. Bower and Gilbody[56] have described a continuum of specialist involvement in primary care mental health with least involvement of secondary care professionals in the education/training model, and increasing involvement through the consultation-liaison model, the collaborative care and the replacement/referral model. In practice, these models are not mutually exclusive. They can complement one another and be adapted to take into account local workforce issues and staff availability.

There are at least four models of working across the interface between primary and secondary care:

1 A **replacement/referral model** is the traditional way in which primary care interfaces with secondary care. In this model, the patient's care is handed over to specialist services by way of a referral, the specialist service only relinquishing care when the patient had been treated. However, over the last 25 years other models have developed.[57] These include:

2 The **consultation-liaison model** allows secondary care professionals to develop ongoing relationships with primary care professionals, not only providing expert advice but also actively liaising with the primary care team, and often attending team meetings or seeing patients jointly with the GP.

3 The **collaborative care model**, as described above has primary care mental health professionals working between primary and secondary care to improve the overall care of patients. These link-workers can provide active follow-up and access to specialist advice and care as needed.[59]

4 In the **training/education model**, secondary care provides education and training to the primary care team which otherwise functions autonomously utilizing secondary care services when needed via a referral system.

How health services negotiate the interface between primary and secondary mental health services is likely to get increasingly complex as secondary care moves away from centralized care through the community mental health teams, to specialist teams such as assessment and brief treatment teams, early intervention teams, and continuing care teams. It is likely that each of these teams will develop its own model of interfacing with primary care. This may allow for a more fluid interface and closer working relationships between primary care and secondary care teams.[59] However, it needs to be managed appropriately or could lead to confusion of roles and responsibilities.

Shared care registers and shared care plans

A **shared care register** is usually a computerized record of all patients jointly cared for by the two services. It might consist of all those who have been discharged from hospital in the past 2 years, all those who have been on a psychotropic drugs for longer than a year, and all psychotic patients known to the GP who have not had an admission to the hospital. The record gives information about the key worker, outpatient clinics are held in the surgeries, and 'good practice protocols' can be developed, so that the case register can be audited against what other teams agree is good clinical practice.

Shared care plans follow on from this development. Such a plan gives the primary care staff information about symptoms which they may expect while the patient is well, likely symptoms in relapse, the name of the key worker, and full details of whom to contact in an emergency both during the day and at night. The plan makes clear who is responsible for medication, and gives an acceptable alternative should the GP find it necessary to vary the medication. It is essential that these plans are mutually agreed between the two teams, rather than being imposed by one team on the other. GPs in England are now being remunerated for keeping registers of patients with psychotic illnesses, dementia, and learning disabilities.

Improving the mental health skills of GPs

About half of GP trainees have a 6-month psychiatry hospital attachment, many of which are considered to be unhelpful for a future generalist career.[59] Many GPs have had no higher professional training in mental health and are not required to do so. There are several GPs who may have previously trained in psychiatry or psychotherapy (e.g. cognitive analytic psychotherapy, CBT, family therapy, etc.) before entering general practice and there is a growing national network of GPs with a special interest in psychiatry (GPsis) with the development of a national course for GPsis to Diploma/Masters level organized by PRIMHE, the National Primary Care Mental Health charity for professionals (www.primhe.org). The GPs network covers most regions and greatly overlaps with the National Trailblazer network (www.iop.kcl.ac.uk). Trailblazers was developed 10 years ago by one of the authors (AT) to bring together professionals from primary and secondary care to work together in pairs to build bridges and

enhance local services by working on a local service development project together with supervision. Trailblazer courses bring pairs together for tutor and peer supervision of projects for up to a year. To date nearly 1000 GPs, psychiatrists, CPNs, PNs, etc., have participated in this way and trailblazer training centres now run in every region in England, supported by the Regional Care Services Improvement Partnership Development Centres (www.csip.org.uk/regions) of the UK Department of Health. International trailblazers is now part of the International Initiative for Mental Health Leaders (www.IIMHL.org) and runs in New Zealand, United States, and England where there is the added interest of comparing service systems by participants who have exchange visits as modules rotate in the three countries. Trailblazers has been positively evaluated for the adult-centred learning approach.[60] Another well-recognized Quality Improvement Programme with a long history of working in primary care has recently focused its methods on common mental health disorders. The Improvement Foundation (www.improvement foundation.org.uk) formerly known as the National Primary Care Development Team (www.npdt.org.uk) have been working with 20 PCTs in England using a Plan Do Study Act cycle (PDSA) to help practices improve their depression care.

As there are few opportunities for primary care workers to obtain mental health skills training, one successful distance learning method involves the use of training DVDs. 'Micro-skills' of assessment or treatment of mental health disorders can be demonstrated by real-life general practitioners with actor–patients in 10 min consultations. The learner is then encouraged to practice these skills using role-plays supplied with the DVDs. A series of existing training materials have been put together for the World Psychiatric Association (WPA) which involves two of the authors (DG and AT) and colleagues from the Institute of Psychiatry, and Prof. Linda Gask and her colleagues at Manchester University. The materials in this WPA package cover depression, somatization, chronic fatigue, schizophrenia, anxiety, and dementia (see www.iop.kcl.ac.uk or www.man.ac.uk for further details).

Summary

At one time, it was asserted that the 'worried well' were treated in primary care, while true mental illnesses were seen by the mental illness services. This was not true when it was asserted, and is even less true now. The great majority of patients with common mental disorders are cared for within primary care, and many of those with severe mental illnesses are only seen in primary care. 'Stepped care' is a model for distributing clinical problems between the services, and 'shared care' refers to the care of patients seen by both primary care and specialist mental health services. Many other workers in primary care now assist GPs with the treatment of mental disorders, and special administrative arrangements within primary care are necessary to ensure that clinical services are available to those with special needs.

In summary, mental disorders in primary care:

- Are an important public health problem
- Frequently present with somatic symptoms
- Are more likely to be detected if the doctor has better communication skills
- Those with disabling physical illnesses are also at greater risk
- Are on average less severe than those seen in specialist care

Further information

Care Services Improvement Partnership (CSIP). Improving Primary care Mental Health Services. National Institute for Mental Health England, Department of Health. 2006. (www.csip.org.uk/resources/publications/primary-care.html) Accessed 22/3/07. A practical guide to developing and improving mental health services in primary care.

http://guidance.nice.org.uk/topic/behavioural. National Institute for Health and Clinical Excellence's website for guidance on mental and behavioural disorders.

Jenkins, R. (ed.) (2004). *WHO guide to mental and neurological health in primary care. A practical guide to the assessment and treatment of mental disorders in primary care* (2nd edn). Royal Society of Medicine Press, London.

References

1. Krueger, R.F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, **56**, 921–6.
2. Goldberg, D.P. and Huxley, P.J. (1992). *Common mental disorders—a bio-social model*. Tavistock, London.
3. Ustun, B. and Sartorius, N. (1995). *Mental disorders in general health care: an international study*. John Wiley, Chichester.
4. Toft, T., Fink, P., Oernboek, E., *et al.* (2005). Mental disorder in primary care: prevalence and co-morbidity among disorders. *Psychological Medicine*, **35**, 1173–84.
5. Verhaak, P., Bensing, J., and Brink-Muinen, A. (in press). Primary mental health care in 10 European countries: patients' demands and GPs' responses. *European Journal of Psychiatry*.
6. Ormel, J. and Tiemens, B. (1995). Recognition and treatment of mental illness in primary care. Towards a better understanding of a multifaceted problem. *General Hospital Psychiatry*, **17**, 160–4.
7. Freeling, P. and Tylee, A. (1992). Depression in general practice. In *Handbook of affective disorders* (ed. E.S. Paykel), pp. 651–6. Churchill Livingstone, Edinburgh.
8. Kessler, D., Bennewith, O., Lewis, G., *et al.* (2002). Detection of depression and anxiety in primary care: follow up study. *British Medical Journal*, **325**, 1016–7.
9. Rost, K., Zhang, M., Fortney, J., *et al.* (1998). Persistently poor outcomes of undetected major depression in primary care. *General Hospital Psychiatry*, **20**, 12–20.
10. Simon, G.E., Goldberg, D., Tiemens, B.G., *et al.* (1999). Outcomes of recognized and unrecognized depression in an international primary care study. *General Hospital Psychiatry*, **21**, 97–105.
11. Dowrick, C.F. (1995). Case or continuum? Analysing GPs ability to detect depression in primary care. *Primary Care Psychiatry*, **1**, 255–7.
12. Wittchen, H.U., Hofler, M., and Meister, W. (2001). Prevalence and recognition of depressive syndromes in German primary care settings: poorly recognized and treated? *International Clinical Psychopharmacology*, **16**, 121–35.
13. Thompson, C., Ostler, K., Peveler, R.C., *et al.* (2001). Dimensional perspective on the recognition of depressive symptoms in primary care: the Hampshire Depression Project 3. *The British Journal of Psychiatry*, **179**, 317–23.
14. Goldberg, D.P., Privett, M., Üstün, B., *et al.* (1998). The effects of detection and treatment on the outcome of major depression in primary care: a naturalistic study in 15 cities. *The British Journal of General Practice*, **48**, 1840–4.
15. Unutzer, J. (2002). Diagnosis and treatment of older adults with depression in primary care. *Biological Psychiatry*, **52**, 285–92.
16. Blanchard, M., Waterreus, A., and Mann, A. (1994). The nature of depression among older people in inner London and the contact with primary care. *The British Journal of Psychiatry*, **164**, 396–402.
17. Crawford, M.J., Prince, M., Menezes, P., *et al.* (1998). The recognition and treatment of depression in older people in primary care. *International Journal of Geriatric Psychiatry*, **13**, 172–6.

18. Unutzer, J., Katon, W., Callahan, C.M., *et al.* (2002). Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *The Journal of the American Medical Association*, **288**, 2836–45.
19. Katon, W.J., Schoenbaum, M., Fan, M.Y., *et al.* (2005). Cost-effectiveness of improving primary care treatment of late-life depression. *Archives of General Psychiatry*, **62**, 1313–20.
20. Hunkeler, E.M., Katon, W., Tang, L., *et al.* (2006). Long term outcomes from the IMPACT randomised trial for depressed elderly patients in primary care. *British Medical Journal*, **332**, 259–63.
21. Audit Commission. (2002). *Forget me not: developing mental health services for older people in England*. Audit Commission, London.
22. Gifford, D.R. and Cummings, J.L. (1999). Evaluating dementia screening tests: methodologic standards to rate their performance. *Neurology*, **52**, 224–7.
23. Woods, R.T., Moniz-Cook, E., Iliffe, S., *et al.* (2003). Dementia: issues in early recognition and intervention in primary care. *Journal of the Royal Society of Medicine*, **96**, 320–4.
24. Turner, S., Iliffe, S., Downs, M., *et al.* (2004). General practitioners' knowledge, confidence and attitudes in the diagnosis and management of dementia. *Age and Ageing*, **33**, 461–7.
25. Downs, M., Turner, S., Bryans, M., *et al.* (2006). Effectiveness of educational interventions in improving detection and management of dementia in primary care. *British Medical Journal*, **332**, 692.
26. Callahan, C.M., Boustani, M.A., Unverzagt, F.W., *et al.* (2006). Effectiveness of collaborative care for older adults with Alzheimer disease in primary care: a randomized controlled trial. *The Journal of the American Medical Association*, **295**, 2148–57.
27. ICD-10. (1993). *The ICD-10 classification of mental & behavioural disorders*. WHO, Geneva.
28. American Psychiatric Association. (1994). *Diagnostic & statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington.
29. WHO. (2001). *Guide to mental health in primary care*. Royal Society of Medicine Press, London.
30. Pingitore, D. and Sansone, R.A. (1998). *Using DSM-IV primary care version: a guide to psychiatric diagnosis in primary care*. *American Family Physician* **58**, 1347–52. Online at: http://www.aafp.org/afp/981015ap/pingitor.html.
31. WONCA. (2005). *International classification committee ICPC--R: International Classification of Primary Care*. OUP, Oxford.
32. The Clinical Terms Version 3 (The Read Codes) NHS Information authority. (2000). Available at http:\\www.coding.nhsia.nhs.uk.
33. Melzer, D., Hale, A., Malik, S., *et al.* (1991). Community care for patients with schizophrenia one year after hospital discharge. *British Medical Journal*, **303**, 1023–6.
34. Kroenke, K., Spitzer, R.L., and Williams, J.B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, **16**, 606–13.
35. Holtz, K., Landis, R.D., Nemes, S., *et al.* (2001). DAPA-PC: development of a computerized screening system to identify substance abuse in primary care. *Journal of Health Quality*, **23**, 34–7.
36. Kobak, K.A., Taylor, L.H., Dottl, S.L., *et al.* (1997). A computer-administered telephone interview to identify mental disorders. *The Journal of the American Medical Association*, **278**, 905–10.
37. Greist, J.H. (1998). Clinical computing: the computer as clinician assistant: assessment made simple. *Psychiatric Services*, **49**, 467–72.
38. National Institute for Health and Clinical Excellence (NICE) (2007) CG23 Depression (amended). *Management of depression in primary and secondary care*. NICE, London. Available from www.nice.org.uk/CG23.
39. Raistrick, H. and Richard, D. (2006). *Designing primary care mental health services: guidebook*. Care Services Improvement Partnership, Department of Health, Leeds.
40. Katon, W., Von Korff, M., Lin, E., *et al.* (1999). Stepped collaborative care for primary care patients with persistent symptoms of depression: a randomized trial. *Archives of General Psychiatry*, **56**, 1109–15.
41. National Institute for Clinical Excellence. (2004). *Anxiety: management of anxiety (panic disorder, with or without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care*. Clinical Guideline 22. National Institute for Clinical Excellence, London.
42. National Institute for Clinical Excellence. (2006). *Obsessive-compulsive disorder: core interventions in the treatment of obsessive-compulsive disorder and body dysmorphic disorder*. National Clinical Practice Guideline Number 31. National Institute for Clinical Excellence, London.
43. National Institute for Clinical Excellence. (2004). *Self harm: short-term treatment and management*. National Clinical Practice Guideline Number 16. National Institute for Clinical Excellence, London.
44. McCrone, P., Knapp, M., Proudfoot, J., *et al.* (2004). Cost-effectiveness of computerized cognitive behaviour therapy for anxiety and depression in primary care randomised controlled trial. *The British Journal of Psychiatry*, **185**, 55–62.
45. Lester, H., Freemantle, N., Wilson, S., *et al.* (2007). Cluster randomised controlled trial of the effectiveness of primary care mental health Workers. *The British Journal of General Practice*, **57**, 196–203.
46. Quality and Outcomes Framework Information available at http://www.ic.nhs.uk/services/qof.
47. Zigmond, A.S. and Snaith, R.P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, **67**, 361–70.
48. Beck, A.T., Mendelson, M., and Mock, J. (1961). Inventory for measuring depression. *Archives of General Psychiatry*, **4**, 561–71.
49. Wilkinson, G. (1992). The role of the practice nurse in the management of depression. *International Review of Psychiatry*, **4**, 311–6.
50. Mann, A.H., Blizard, R., Murray, J., *et al.* (1998). An evaluation of practice nurses working with general practitioners to treat people with depression. *The British Journal of General Practice*, **48**, 875–9.
51. Mynors-Wallis, L.M., Gath, D.H., Lloyd-Thomas, A.R., *et al.* (1995). Randomised controlled trial comparing problem-solving treatment with amitriptyline and placebo for major depression in primary care. *British Medical Journal*, **310**, 441–5.
52. Holden, J.M., Sagovsky, R., and Cox, J.L. (1989). Counselling in a general practice setting: controlled study of health visitor intervention in treatment of postnatal depression. *British Medical Journal*, **298**, 223–6.
53. Appleby, L., Warner, R., Whitton, A., *et al.* (1997). A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression. *British Medical Journal*, **134**, 932–6.
54. Mellor-Clark, J., Simms-Ellis, R., and Burton, M. (2001). *National survey of counsellors in primary care: evidence for growing professionalisation*. Royal College of General Practitioners, London.
55. Bower, P. and Rowland, N. (2006). Effectiveness and cost effectiveness of counselling in primary care.[update of Cochrane Database of Systematic Reviews, 2002; (1): CD001025; PMID: 11869583]. *Cochrane Database of Systematic Reviews*, **3**, CD001025.
56. Bower, P. and Gilbody, S. (2005). Managing common mental health disorders in primary care: conceptual models and evidence base. *British Medical Journal*, **330**, 839–42.
57. Gask, L., Sibbald, B., and Creed, F. (1997). Evaluating models of working at the interface between mental health services and primary care. *The British Journal of Psychiatry*, **170**, 6–11.
58. Von Korff, M. and Goldberg D. (2001). Improving outcomes in depression: the whole process of care needs to be enhanced. *British Medical Journal*, **323**, 948–9.
59. Turton, P., Tylee, A., and Kerry, S. (1995). Mental health training needs in general practice. *Primary Care Psychiatry*, **1**, 197–9.
60. Brown, C., Wakefield, S., Bullock, A., *et al.* (2003). A qualitative evaluation of the Trailblazers teaching the teachers programme in mental health. *Learning in Health and Social care*, **2**, 74–82.

7.9

The role of the voluntary sector

Vanessa Pinfold and Mary Teasdale

What is the mental health voluntary sector?

The voluntary sector plays an important role in the mental health field across the world. Originally set up and run by volunteers, the prime motivation and purpose for this sector is improving the lives of people affected by mental health problems by doing things differently, tirelessly pushing for change, never giving up hope and working alongside service users (also known as patients and consumers) and their families every step of the way. In some countries it has been labelled the 'third sector' to distinguish it from other organizational sectors namely industry (private sector) and government (public or statutory sector).

The voluntary sector is not, however, a cohesive group of organizations and across the mental health community each one operates with its own specific remit. Some of these organizations have become large businesses providing a wide range of services under contract with statutory agencies. Others choose to avoid employing staff and are still run entirely by committed volunteers. Many rely on voluntary donations in order to remain fiercely independent of government. Each has its own aims and mission, core stakeholder group, trustee and membership structures, management systems, governance procedures, and a unique portfolio of activities.

Some mental health organizations focus activities on mental health or emotional well-being specifically (e.g. Finnish Association for Mental Health). Others target a social problem and support all those affected such as charities working with the homeless, refugees, victims of domestic violence, or young offenders, including people with mental health problems. There are organizations that primarily campaign, educate, advocate, lobby, and promote self-help resources such as EUFAMI—an association for families across Europe and SANE Australia. In parts of the world, including Eastern Europe, there are particular challenges in mental health resulting from poverty, dislocation of the population, and insufficient resources for health. Some states, like Armenia, have no mental health services. Here, the Catholic Agency for Overseas Development (CAFOD) is working with Armenia's Association of Child Psychiatrists and Psychologists to provide therapy and mental health care and to increase awareness and understanding in order to overcome prejudice.

Wherever they work, key characteristics of voluntary sector organizations include an independent position, a strong values base, empowerment principles, non-profit distributing of resources, passionate commitment to the work focus, rooted to service user, and carer experiences and they are always striving for changes to provide people with mental health problems better provision and opportunities. Most started as local support groups but some have grown into large national organizations with considerable political leverage. The National Schizophrenia Fellowship (known today as Rethink) was founded in 1972 by a group of families concerned that relatives of people with schizophrenia had no support for themselves. In this chapter we draw on the Rethink experience to illustrate how the voluntary sector contributes to, and shapes, modern psychiatry. Although the English experience does not directly map onto those in other countries across the world, there are similarities and we seek to highlight these through the use of international examples where possible.

Rethink severe mental illness

Rethink is a membership charity with 7500 members (service users, carers, mental health professionals, the general public) whose mission is 'to support everyone affected by severe mental illness recover a better quality of life'. It adopts a recovery-orientated approach to supporting the individual and their family through periods of ill health and their journey of recovery. This perspective is significantly different from that of clinicians and statutory providers, who have not been through the experiences common to people who suddenly have to cope with severe mental illness. Rethink staff descriptions of their role include:

> *Bridging: linking a person with non-judgemental delivery of services connected with service user and carer experience*
>
> *Ensuring service users are heard and needs met more holistically*

Initially providing mutual support, the organization later offered information resources which address the problems commonly encountered by service users and carers, like difficulty in gaining access to services or funding for appropriate care. The emphasis is on finding successful strategies which achieve solutions. Advocacy is provided for individuals and families whose needs are not being met. The experiences of service users struggling to cope provide detailed evidence which is used to develop Rethink's policy on the mental health issues which reach the political arena and also as the basis of campaigns on stigma and discrimination. Research and

surveys of service user and carer views form the basis of reports on vital issues, like how information can be provided to carers with due respect for the service user's privacy and autonomy. Guidance on good practice may be developed and sometimes training for professionals. Media activity has publicized Rethink's campaigns and using the internet has made dissemination of information cheaper and easier than it used to be.

Rethink's activities also include the provision of front-line services in partnership with the National Health Service and Social Services (statutory sector) for example supported housing schemes, advocacy projects, community resource centres, carer support services, employment and training programmes, school education projects, and mentoring programmes for young people. In 2007, Rethink ran 350 front-line mental health services and employed approximately 1300 staff.

Does the voluntary sector make a difference?

The voluntary sector makes a substantial contribution to both the image of psychiatry and its practices. For example in New Zealand, the Mental Health Foundation has worked in partnership with the Ministry of Health and other agencies to run a successful anti-discrimination campaign—like minds, like mine (whakaitia te whakawhiu i te tangata) for the past 10 years. This is an internally renowned mental health awareness programme that is transforming how the New Zealand public engages with mental health issues. Non-Governmental Organizations (NGOs) can also bring influence to bear at a national level and bring people together to plan new service models. The World Fellowship for Schizophrenia and Applied Disorders (WFSAD) supported a workshop in East Africa in 2003 where service users and families could meet with government ministers and medical professionals to discuss plans for achieving effective health care delivery. In England, the voluntary sector has collectively ensured that the published clinical guidelines describing best practice for the treatment of schizophrenia in 2002 took note of service user and carer treatment preferences. In India, Action for Mental Illness (ACMI), an advocacy initiative has achieved tax concessions for those with mental illness and their carers and also maintenance allowances equivalent to those provided to people with physical disabilities.

Standards of mental health care delivery vary dramatically—with cases of human rights abuses in psychiatry being documented in some countries and innovative services emerging in others in response to local demands. The voluntary sector can, and does, bring the spotlight on both ends of the service delivery spectrum (good and bad) and demand better for everyone. It also leads the way by developing innovative solutions both in terms of service delivery, public education, and self-management techniques. In the United States, NAMIs Peer-to-Peer Education Course is a 9-week experiential education course on the topic of recovery for anyone with serious mental illness who is interested in establishing and maintaining wellness. In Canada, a family to family network was established for first episode psychosis families by the Canadian Mental Health Association. The recovery model is being pioneered and embraced by the voluntary sector across the world. However, this relatively new approach requires a change of attitude by both service users and professionals as shown by the Scottish recovery network programme.

Common themes tend to emerge through the campaigns of voluntary organizations across the world, in spite of differences in wealth and varying stages of development. In most places the demands are for earlier intervention, better crisis response, more support for families, and less use of physical restraint on the ward when coping with challenging behaviour. The transition to care in the community presented a new set of problems, not least the need for different professional skills and adequate resources. And in the later part of the last century, the controversy over community treatment orders was raging in many countries. Changes in government policy and concerns over access to newer treatments resulted in new alliances in many countries, and the voluntary sector, clinicians and other professionals often learned to work together, recognizing how in alliance they could lobby more effectively for better law and more resources.

An excellent example of successful partnership working in England and Wales has been the formation of a coalition known as the Mental Health Alliance involving 80 organizations. Professional bodies have joined, including the Royal College of Psychiatrists as prominent and active members. The Mental Health Alliance has opposed the government's proposals for reforming the Mental Health Act 1983 for the past 8 years and has managed to achieve some change in the content of the legislation as well as delaying the whole process.

The role of critical friends

The independence of organizations is an essential characteristic of the voluntary sector. These bodies do occupy the territory of 'critical friends' to both the statutory and private sectors of the mental health community, monitoring activity and speaking out in praise of positive developments but also highlighting when things are wrong.

For example at Rethink experience shows that misdiagnosis can result in inappropriate care and treatment with tragic consequences like imprisonment, suicide, or even homicide. Therefore Rethink supports service users in obtaining expert second opinions. Providing families or individual service users with accurate but understandable information enables them to challenge the opinion of a psychiatrist or medical team if this seems appropriate.

Rethink also advocates for families by using the complaints procedures or by providing legal representation at inquest hearings in order to draw attention to deficiencies in support, care, and treatment. They focus efforts on cases where systemic problems played a part, like a poor approach to risk assessment or refusal to accept information from families. The aim is to persuade the Ombudsmen or coroners to recommend improvements in the local policies and procedures in order to improve the quality of services.

In recent times there has been a spotlight on mental health services' engagement with people from Black and Minority Ethnic groups (BME) in England. The BME voluntary mental health organizations have formed a network which aims to reduce inequalities and promote good practice in mental health for racialized groups. The Network has been very critical of the Government's proposals to amend mental health legislation and has criticized the statutory health sector for failing to meet legal requirements on race equality. Similar issues arise in the United States where the National Council of La Raza (NCLR)—the largest national Hispanic civil rights and advocacy organization in the United States—works to improve

opportunities for Hispanic Americans. They report that Latinos are at a disproportionately high risk for depression and other conditions associated with mental illness, and are also much less likely to seek treatment or receive quality culturally and linguistically competent care.

Conclusion

The voluntary sector is a dynamic and vital part of any mental health system. Rooted in the experiences of mental health service users and carers, voluntary sector organizations across the world ensure that the voices of 'experts by experience' directly influence campaigns, policy debates, service redesign, and project planning and treatment guidelines. The sector is, however, fragile and in some countries organizations are increasingly dependent on state funding which could undermine their autonomy and independence. Psychiatrists can support their local voluntary organizations by joining them—as members, as campaigners, and as educators. The sector can also support psychiatrists, helping to transform the public image of psychiatry and encouraging young people to take an interest in mental health as a career option. The alliances forged with psychiatrists and their representative bodies are crucial for improving the quality of mental health services and to effectively tackle stigma and discrimination. We do need each other in order to deliver better outcomes for mental health service users and their families.

Further information

The websites for some of the voluntary organizations referenced in the chapter are:

Rethink: www.rethink.org

Canadian Association for Mental Health: www.camh.ca

EUFAMI—European Federation of Associations of Families of People with Mental Illness: www.eufami.org

Mental Health Foundation in New Zealand: www.mentalhealth.org.nz

National Alliance on Mental Illness: www.nami.org

SANE Australia: www.sane.org

Scottish recovery network: www.scottishrecovery.net

The World Fellowship for Schizophrenia and Allied Disorders (WFSAD): www.world-schizophrenia.org

7.10

Special problems

Contents

7.10.1 The special psychiatric problems of refugees

Richard F. Mollica, Melissa A. Culhane, and Daniel H. Hovelson

While the forced displacement of people from their homes has been described since ancient times, the past half-century has witnessed an expansion in the size of refugee populations of extraordinary numbers.[1,2] In 1970, for example, there were only 2.5 million refugees receiving international protection, primarily through the United Nations High Commission for Refugees (**UNHCR**). By 2006, UNHCR was legally responsible for 8.4 million refugees. In addition, it is conservatively estimated that an additional 23.7 million people are displaced within the borders of their own countries. Although similar in characteristics to refugees who have crossed international borders, internally displaced persons do not receive the same protection of international law. Adding all refugee-type persons together, the world is forced to acknowledge the reality that over the past decade more than 10 000 people per day became refugees or internally displaced persons.

The sheer magnitude of the global refugee crisis, the resettlement of large numbers of refugees in modern industrial nations such as Canada, the United States, Europe, and Australia, and the increased media attention to civil and ethnic conflict throughout the world has contributed to the medical and mental health issues of refugees becoming an issue of global concern. This chapter will focus on a comprehensive overview of the psychiatric evaluation and treatment of refugees and refugee communities. Although this mental health specialty is in its infancy, many scientific advances have been made that can facilitate the successful psychiatric care of refugee patients.

Definition

The definition of a refugee as outlined in the 1951 Convention and 1967 Protocol relating to the Status of Refugees is presented in Box 7.10.1.1.[3] A person or persons who has passed over from one country into another seeking protection from violence and who cannot return to his country of origin because of fear of persecution or injury is considered a refugee according to international law.

The 1951 Convention Relating to the Status of Refugees was drawn up by the United Nations parallel to the creation of UNHCR. This Convention and the subsequent 1967 Protocol establishes

Box 7.10.1.1 Definations according to the 1951 convention and 1967 protocol relating to the status of refugees

Article 1—Definition of the term 'refugee' A(2) [Any person who] . . . owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having nationality and being outside the country of his former habitual residence . . . , is unable or, owing to such fear, is unwilling to return to it. (As amended by Article 1(2) of the 1967 Protocol.)

Article 33—Prohibition of expulsion or return (refoulement)
(1) No contracting state shall expel or return (refouler) a refugee in any manner whatsoever to the frontiers of territories where his life or freedom would be threatened on account of his race, religion, nationality, membership of a particular social group or political opinion.

international law for the definition of refugees as well as the protection accorded to them. It also articulates the important principle of *non-refoulement* (Box 7.10.1.1), which states that no refugee can be returned to his or her country of origin or any other location where there is any probability that he or she will be harmed. These legal definitions indicate that a refugee is not an economic migrant or a traditional immigrant. Sadruddin Aga Khan, in a seminal report, was one of the first High Commissioners to acknowledge the human rights violations that are primarily responsible for the generation of refugee populations.[4] Corresponding to these international covenants, the international community has focused on the protection of refugees. There are two components to protection that are classically viewed by UNHCR as an essential aspect of its mandate. These two elements include protection against:

1 ongoing violence and potential injury to the refugee including being denied proper asylum and involuntary repatriation;

2 lack of adequate food, water, clothing, and other forms of material assistance.

The UN Declaration of Human Rights adopted by the United Nations in December 1948 and the United Nations Convention against Torture and Other Cruel Inhuman or Degrading Treatment or Punishment adopted in December 1984 extends the basic principles of refugee protection and asylum, and reaffirms the principle of *non-refoulement*. In most refugee crises, not withstanding the political and military barriers to protection, UNHCR and the international community strive to offer refugees a safe asylum and basic humanitarian aid.

Trauma and torture

By definition, most refugees have experienced traumatic life events of extraordinary brutality. Since the Second World War, empirical studies have investigated the relationship between mass violence, the refugee experience, and psychiatric morbidity. The earliest research focused on survivors of the Nazi concentration camps.[5–8] Shortly after the Second World War, Eitinger and his colleagues gave a detailed account of their medical and psychiatric examinations of concentration camp survivors. They postulated that the traumatizing process had a dual nature. They described the somatic traumas of captivity, such as head injury, hunger, and infections, as leading to a 'psycho-organic syndrome', and the predominately psychological traumas as leading to other psychiatric disorders such as depression. Thygesan's studies of concentration camp survivors in Denmark revealed similar results.[9,10] These early pioneering investigations of the psychosocial sequelae of the Nazi concentration camps established a preliminary baseline of traumatic outcomes for future generations of refugees, many of whom had experienced the trauma of similar experiences in Cambodia, Bosnia-Herzegovina, and elsewhere.

Increasingly, civilian populations carry the burden of ethnic conflict and mass violence. It is now estimated that more than 80 per cent of casualties caused by the recent violence in Africa, Asia, and Europe have primarily affected non-combatants.[11] Extensive research has revealed the major trauma events experienced by refugee populations fall into the eight groups below:

1 material deprivation

2 war-like conditions

3 bodily injury

4 forced confinement and coercion

5 forced to harm others

6 disappearance, death, or injury of loved ones

7 witnessing violence to others

8 brain injury

Every refugee situation will have a range of traumatic events that will fall into each of these categories that are unique or characteristic of a specific conflict. It is essential that the specific types of violence experienced by a given refugee population are well known to the psychiatric clinician who can use this knowledge to assess potential traumatic outcomes.[12] In addition to many unique forms of violence occurring in different refugee settings, the meaning of violent events also differs across cultures. Anecdotal, clinical, and epidemiological evidence suggests that certain categories of refugee trauma are more potent than others in producing psychiatric morbidity and other traumatic outcomes. Brain injury, sexual violence, torture and other forms of bodily injury, coercion, and forced confinement have great potential of causing psychiatric harm in refugees exposed to these events. Consistent with indicators of the 'potency' of specific trauma events, there evidence of a dose–effect relationship between cumulative trauma and psychiatric symptoms.[13] The personal aspects of human suffering associated with specific types of trauma, such as the murder of a child or the disappearance of a family member are still relatively undefined but obviously very difficult.

Many refugees have actually experienced torture. Recent research states that the most significant finding in the last 7 years may be that either torture has become more prevalent worldwide or the total number of events reported has increased, likely as a result of advocacy and elevated media attention.[14] After 25 years of research on treatment work with torture survivors, still no consensus exists for effective interventions within the field.

Conceptual model of traumatic outcomes

Research on refugees has revealed the persistence of negative health and social outcomes decades after their initial experience of violence and dislocation. Emergence of standardized criteria for psychiatric diagnoses and disability and the demonstrated ability to elicit trauma events through simple screening instruments in culturally diverse populations have allowed evidence to accumulate suggesting a model of traumatic outcomes associated with the refugee experience. This model is primarily based upon the classic epidemiological triad which describes the interaction between host (i.e. the refugee), agent (i.e. traumatic life experiences), and environment (e.g. refugee camp) in the pathogenesis of psychiatric disorders.[15,16] This model, illustrated in Fig. 7.10.1.1, allows equal attention to be given to all aspects of the refugee experience.

The model in Fig. 7.10.1.1 has three major elements. First, it suggests that the major medical outcomes associated with the refugee experience are medical illness, psychiatric disorders, and disability. Second, trauma and the personal and environmental characteristics of the refugee describe the major risk factors associated with violent outcomes. Third, the direction of the causal arrows in the model do not imply a lack of reciprocal relationships where none is indicated; instead they indicate what most investigations consider

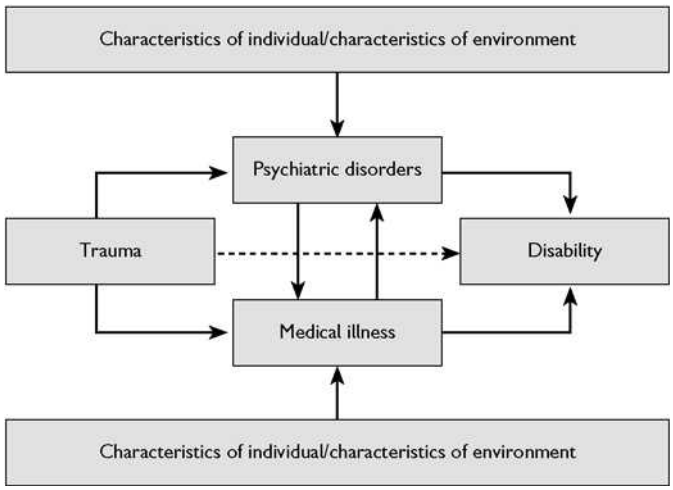


Fig. 7.10.1.1 Conceptual model of refugee risk factors and traumatic outcomes.

to be the most dominant causal relationship. Despite its limitations, this simple conceptual model can provide the psychiatric professional with a scheme for approaching the refugee patient from either a clinical or public health perspective. The importance of the socio-cultural and political context unique to each refugee situation and its impact on each of the model's pathways cannot be overstated, since refugees come from diverse cultural groups and political experiences.

Health status and medical illness

Refugees experience many diseases and chronic debilitating conditions, such as starvation and landmine injuries that have both immediate and long-term effects on their physical health. Every refugee situation involves many unique but common acts of violence leading to major medical sequelae. For example, gender-based violence and rape, which were instruments of ethnic cleansing in the Balkans, resulted in pregnancy, medically complicated self-administered abortions, and sexually transmitted diseases.[17] Extensive documentation of refugee survivors of torture describes the medical sequelae of torture and the causal link between refugee trauma and medical mortality and morbidity over time.

Throughout the process of migration, refugees are frequently exposed to infectious diseases and physical or psychological trauma that can have a profound impact on their overall health and well-being. Refugees are at increased risk for diseases such as obesity, diabetes, and heart disease, and often encounter significant barriers to employment and health care when attempting resettle. For these reasons, programmes implemented in the refugees' country of resettlement designed to improve health-seeking behaviours and overall self-care can have a positive, lasting impact on refugees' physical and mental health.[18]

Psychiatric symptoms and illness

Observations since Kinzie *et al.*[19] and Mollica *et al.*[20] first diagnosed PTSD in Cambodian refugees, have made the cultural validity of PTSD seem almost certain. However, this reality does not negate the importance of culture-specific symptoms related to trauma that are independent of PTSD criteria. Recent large-scale epidemiological studies of refugee populations have confirmed the high prevalence of major depression and PTSD in Western (e.g. Bosnian[21]) and non-Western (e.g. Cambodian[22] and Bhutanese[23]) refugee communities. The mental health impact of major depression, which presents both as a comorbid disorder with PTSD and alone, is chronic, severely disabling and demands the attention of the clinician working with refugees. Longitudinal data indicates that 45 per cent of Bosnian refugees who met criteria for PTSD, depression or both continued to meet criteria for these disorders 3 years later.[24] Similar results were found in a longitudinal study of Cambodian refugees 20 years after resettling in the United States.[25] As with Western populations, depression in refugees tends to be under-diagnosed and can be expressed as somatic complaints. A study of Vietnamese refugees showed high prevalence based on self-report, but high rate of physician under-diagnosis. Most patients with depression (95 per cent) presented with physical complaints.[26] These finding underscore the importance of depression screening especially in the primary care setting.

New research indicates that there may be memory problems in refugees with PTSD. When asked to recall traumatic events refugees with PTSD reported an increased number of traumatic or torture events over their baseline report, as compared to those with other psychiatric disorders who showed no change or decreases in number of events reported.[27] Substance use disorders are often overlooked in refugee populations but are often comorbid with PTSD.[28] Early in the 1980s reports of Hmong refugees using opium emerged.[29] Substance use disorders have been shown to have a delayed presentation of 5 to 10 years after the initial settlement of the refugee.[30] However, substance use disorders may vary by population. One recent study of Cambodian refugees in the United States reports low rates of alcohol use in the past 30 days.[31] Screening for substance use disorders especially in primary care is important. Complex grief reaction and chronic insomnia are also prevalent in this population.

(a) Head injury

Clinical evidence is also emerging identifying head injury as a cause of significant psychopathology in refugee survivors. Head trauma is one of the most common forms of torture, so much so that reports of torture almost always imply that some kind of head trauma occurred. Many head injury survivors experience seizures and headaches, as well as behavioural disturbances such as aggressiveness, irritability, and sleep disturbances. Recent research on Vietnamese ex-political detainees who experienced head trauma while in captivity indicates that the number of head injuries is related to a decrease in executive functioning, those with head injury have increased risk of developing PTSD, and had decreased cortical thickness in several brain regions (Mollica *et al.* 2007, unpublished). The long-lasting effects of head trauma are serious and pervasive, affecting both the injured individual and their family. The presence of head trauma in torture victims has been overlooked in past research, but awareness is growing.

(b) Functional status and disability

Although the functional status of refugees at the emergency and long-term ends of the continuum of the refugee experience has received little attention, the significance of this traumatic outcome is beginning to emerge. Until recently, the standard operating model of refugee protection has not been asked to determine the long-term socio-economic damage caused by the refugee experience. The answer to this question is extremely important to the

recovery of societies in which the majority of the population has been displaced. In many societies, the refugee experience is the majority experience, which has strong implications for future socio-economic development. While the prevalence of functional impairment and disability is unknown in refugee populations, a recent epidemiological study of Bosnian refugees in Croatia reveals that functional disability may, in fact, be extremely high, especially in elderly refugees.[21] Furthermore, disability may be exacerbated in refugee survivors who have both chronic medical and psychiatric disorders.

Psychiatric assessment

This section reviews key factors unique to the psychiatric evaluation and diagnosis of the refugee patient.

Primary care: the proper setting for a refugee clinic

The psychiatric literature has generally stressed the importance of evaluating and treating refugee patients in a primary health care setting, whether in a refugee camp or in a country of resettlement. Four factors seem to support this viewpoint:

1 refugee patients seldom self-refer to psychiatry;

2 in many societies considerable stigma is associated with psychiatry but not with primary care medicine;

3 the majority of refugees seek out the care of their local medical doctors and indigenous traditional healers for the relief of their emotional suffering;

4 most refugees have associated medical and psychiatric disorders.

Considerable field experience has shown that establishing a mental health programme within a health facility where refugees already seek medical care can result in the highly successful utilization of psychiatric professionals and treatment.

Cross-cultural psychiatric assessment and diagnosis

Early research describing the psychiatric status of refugee survivors, especially those who had been tortured, refrained from the use of psychiatric diagnoses because of a prevailing perception that the observed symptoms were a normal response to horrific life experiences.[12] Similarly, many medical anthropologists believed that Western psychiatric diagnostic classifications were not relevant to the assessment of suffering in non-Western populations.[32] Despite these reservations, the emergence of standardized diagnostic criteria for major depression and post-traumatic stress disorder (PTSD) have allowed for the cultural validity of these diagnoses to be tested in a number of refugee settings. Cross-cultural research suggests that assessments of psychiatric illness should begin with phenomenological descriptions of folk diagnoses or culture-specific syndromes.[33] Important methods of exploring the validity of DSM-IV diagnoses in cross-cultural settings have included using culturally valid definitions of functioning and mental health problems based on local views of maladaptive thoughts and behaviour in response to distress, not preconceived Western diagnostic categories[34]; however, to date, not a single culture-specific illness associated with the mass violence and torture experienced by refugees has been defined.[35] On the contrary, the criteria for the two major diagnoses associated with violence in Western society, i.e. major depression and PTSD, have been successfully applied to refugees from many parts of the world. While high rates of PTSD can be measured in refugee patients and traumatized civilians, it is not known if other culture-specific symptoms not part of the DSM-IV criteria that may have greater clinical relevance and meaning to a specific refugee group. A general principle demonstrated by the World Health Organization cross-cultural study of depression,[36] i.e. that while some depressive symptoms may be present across cultures, they may not be the symptoms most strongly endorsed by the patient; this principle may also apply to PTSD. Figure 7.10.1.2 provides an illustration, which can help to address the problem of psychiatric diagnoses in refugee patients. This figure suggests that, until further research is forthcoming, the psychiatric provider needs to determine clinically whether the refugee patient is presenting with scenario A, B, or C.

The high prevalence of psychiatric symptoms associated with trauma in refugee populations neither affirms nor negates the 'normalization' of these symptoms. A narrow medical viewpoint could create a psychiatric redefinition of refugee mental health problems that would place the majority of refugees in a 'mentally ill' box without any access to individual psychiatric care; on the other hand, the hostility of many humanitarian aid workers toward psychiatry has denied the seriously mentally ill refugee legitimate access to psychiatric treatment. There will probably be a compromise at the intersection of public health objectives and the protection goals of humanitarian aid workers. In future, the presence of chronic and severe disability in refugee survivors will be the gold standard which drives the psychiatric and humanitarian rehabilitation of refugee survivors.[37,38]

Psychiatric screening

If mental health practitioners are to treat refugee patients, they must be able to assess the refugee's major risk factors and traumatic outcomes. Simple screening instruments culturally adapted to the language, trauma, and symptoms of refugee patients have been found to be extremely effective as well as being well received by refugees themselves.[39–41] For example, a simple well-known

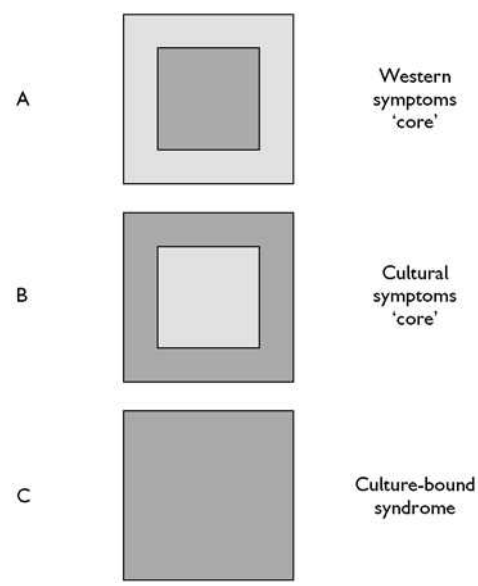


Fig. 7.10.1.2 Comparison of diagnostic classifications.

screening instrument adapted from the Hopkins Symptom Checklist allowed Indochinese patients for the first time to provide their symptoms with little distress. The development of the Harvard Trauma Questionnaire further revealed that these same patients could provide answers to lists of possible trauma events and symptoms without becoming retraumatized. Overall, two major lessons have been learned from the use of simple checklists with refugee survivors.

1 The checklist acknowledges the traumatic life experiences of the refugee survivors and *de facto* gives them permission to elaborate on the details of their trauma.

2 The checklist, as a simple medical test, helps refugees to 'put words around' events and symptoms that would be too emotionally overwhelming for them in an open-ended interview.

Extensive research regarding appropriate methods of development, translation, and validation of the clinical screening instruments such as the Hopkins Symptom Checklist and Harvard Trauma Questionnaire has been widely disseminated.[29,42]

Understanding somatic complaints

For years it was considered standard psychiatric wisdom that emotional distress, especially in non-Western patients, was primarily somatic in character; and that, because of their dominant expression of suffering through somatization, these patients were not psychologically minded, making them incapable of participating in psychotherapy or counselling. Extensive clinical experience and scientific studies with refugee patients throughout the world have led to a revision of these previously dominant professional attitudes. While it may be valid that some refugees primarily present their emotional suffering through somatic complaints, it has also been found that it is fairly easy for the medical practitioner to obtain from refugee patients deeper insights into their feelings and their beliefs as to the causes of their emotional distress. Often because of severe victimization, refugees as a group will not readily share their experiences of trauma and related medical and psychiatric injuries unless they are in a highly confidential environment where it is clear that the medical team can be trusted. In fact, the ability of refugee patients to extend themselves beyond their initial somatic complaints, as well as participating in clinical dialogue with the psychiatric practitioner as to the nature of their mental distress, is more the rule than the exception and is a major goal of assessment.[43]

Psychiatric treatment

The trauma story

The trauma story emerges as the centrepiece of any treatment approach. Every refugee patient has at least one traumatic experience that figures prominently in his or her life history. The trauma story is a living reality and is present for every patient. The trauma story is also present for the clinician. Yet, the story can be elusive and difficult for both patient and doctor to share. Often times when the patient is ready to tell the trauma story, the clinician is not ready to hear it. More frequently, when the clinician is ready for the patient to tell his or her story, the patient is unwilling. Therefore, one of the major goals of psychotherapy with refugees is to allow the trauma story to emerge gently and become a familiar and acceptable theme, which the refugee is not shamefully hiding from his or her therapist and ultimately his or her family and community.[44]

(a) Debriefing

Although commonly conducted within populations who have experienced trauma, typically by relief workers or crisis counsellors, debriefing has been shown to have negative effects on psychiatric outcomes and should be avoided at all costs.[45] As described above, the trauma story should be allowed to emerge naturally and at the patient's own pace.

(b) Treatment

Psychiatric treatment of refugee patients should primarily be based upon standard psychiatric practices for mentally ill patients, including the appropriate use of psychotropic medicines. However, a number of novel therapeutic approaches have been used with refugees. Recent research in refugee populations has indicated that cognitive behavioural therapy (CBT) is efficacious for treatment-resistant PTSD and panic attacks[46,47] and in refugees who have experienced torture.[48] Interpersonal psychotherapy for depression has been found to be useful in rural Uganda.[49] This approach may also be effective with resettled refugees in Western countries. In female veterans of war, prolonged exposure therapy was more effective over time than other interventions for reducing symptoms of PTSD in female veterans of war. It may be possible and beneficial to use prolonged exposure therapy for others with PTSD.[50] Similarly, although no research has specifically examined the effects of therapies such as eye movement desensitization and reprocessing (EMDR) for refugees, positive results have been seen in a variety of different studies done with traumatized populations, including individuals with PTSD. Traditional medicine, which includes diverse health practices, knowledge, and beliefs is widely accepted and practiced worldwide. Traditional medicine typically relies upon local classifications of emotional distress and its treatment can incorporate plant-, animal- or mineral-based medicines, as well as spiritual therapies or massage.[45] Traditional healing approaches were widely used for the Cambodian refugee crisis of the 1990s.[51–53] Although widely used, no conclusive research has yet been done using complementary treatments such as massage, acupuncture or herbal medicine in the treatment of torture survivors.[14] Recent research with Cambodian refugees it was shown that interventions supporting work, altruism, and spirituality for refugees might serve as protective factor for the onset of psychiatric disability.[54]

Special considerations

A number of special considerations have emerged from the many therapeutic approaches that have been tried with refugees.

Family focus

Treatment should be directed at the entire family of the identified refugee patient. The refugee crisis tends to initiate a process in which surviving family members become each other's major social support system. The disruption and disintegration of stable communities and traditional social supports place the entire burden of survival upon family members. Sometimes in extreme situations that affect the family unit such as when a refugee is experiencing symptoms of depression, psychosis, or social withdrawal, the psychiatrist must reach out to the patient and establish a one-to-one therapeutic relationship until family members can be of assistance.

Cultural sensitivity

The cultural sensitivity of refugee mental health practitioners is essential to the proper therapeutic relationship to the refugee patient. This means that the refugee mental health practitioner must be informed as to the type of trauma experienced by the refugee and its socio-cultural meaning, the cultural idioms by which human suffering is expressed in a given community, and the social stigma associated with mental illness. Despite the refugees' own adversity, long-standing social prejudices against mental illness will persist throughout the refugee crisis.

Little cross-cultural literature exists on the relationship between refugee patient, Western professional, and bilingual interpreter. The use of refugees as interpreters can be problematic. Ideally, health professionals from the refugee communities should be recruited to participate in psychiatric intervention. These individuals have had professional medical training as to the importance of confidentiality and can also provide insight into the cultural nuances of the doctor–patient relationship. The use of untrained interpreters from the refugee community should be avoided if at all possible. This is especially true of family members or members of a community or government institution that has previously threatened the security of the refugees. In addition, medical practitioners must respect patient dignity by not allowing young people to interview community elders or males from the community to ask refugee women explicit sexual questions and/or witness or be exposed to an undressed refugee woman during a medical examination. The more the bicultural interpreter can function in the role of a trained mental health paraprofessional, the more successful will be the therapeutic experience of all involved.

Caring for the seriously mentally ill

Despite numerous biases against psychiatric intervention in refugee populations, psychiatrists are well positioned to treat the seriously mentally ill refugee. In most refugee camps, those in immediate danger are the psychotic and depressed refugees with suicidal ideation. These individuals have difficulty coping with their ongoing crisis and have high mortality rates. Once resettled, these refugees with serious mental health issues need appropriate medical and mental health services.[55]

Special treatment needs of gender-based violence

Many colleagues in the field of refugee mental health have sought to break the 'conspiracy of silence related to sexual violence'.[32] While this issue has been well documented in refugee communities, it was not until the Bosnian conflict that rape was finally accepted by the international community not as a criminal act but as a crime against humanity. This recognition of gender-based violence as the most common type of torture of women, as well as a major terrorist instrument of war, has contributed to the protection and psychiatric care of sexually abused refugee survivors. While the political will now exists to condemn rape, and to protect refugees from it, the cultural stigma and corresponding social punishments of women considered 'tainted' by sexual violence, sometimes including their murder ('honour killing') by relatives, continues to make the psychiatric care of these survivors extremely difficult. In most cultures, rape remains a secret issue caused by the refugee's extreme resistance to reveal any details to the physician. Psychiatric practitioners must approach this issue with extreme caution in a strictly confidential manner. They must also be aware of the severe consequences to the patient if the rape experience becomes public knowledge, even to the patient's family members.

Risk and resiliency factors

The pre- and post-conflict personality characteristics of refugees that increase resiliency and reduce psychiatric distress and disability are not known. While certain demographic characteristics have been associated with negative traumatic outcomes, these characteristics may be confounded by other risk factors. Women, especially widows, seem especially vulnerable to negative refugee effects. UNICEF has extensively reviewed those risk factors associated with the vulnerability of refugee children and adolescents.[56] Data from Bosnian refugees correspondingly reveal the high rates of disability associated with trauma and psychiatric comorbidity in the elderly.

Lessons learned from studies of political prisoners in Turkey reveal the importance of a well-established political world view as a major protection against the long-term human suffering associated with torture.[57] Studies of Bhutanese refugees in Nepal confirm the possible protective function of Buddhism in devout refugee practitioners of this religion.[23] Anecdotal reports by refugees themselves consistently confirm the emotional safety that they have found in their spiritual and religious beliefs and practices. Finally, recent research findings concur with the earlier research in concentration camp survivors and prisoners of war that prior psychiatric history and premorbid personality factors may have little effect on the psychiatric sequelae of traumatic refugee experiences. In resettlement countries, opportunities related to family unification, learning to speak the new country's language, and employment have clearly been shown to be associated with decreased psychiatric morbidity over time. Those who provide psychiatric care for refugees must keep in mind the environmental opportunities that can increase their overall resiliency.

Reducing risk and maximizing resiliency

Figure 7.10.1.1 provides a readily accessible model for psychiatric practitioners to determine how they can reduce the risk factors associated with psychiatric disorders as well as promote the resiliency of the refugee patient. The section on risk factors provides many useful insights into the importance of enhancing the refugee's active role in work, spiritual participation, involvement in altruistic behaviour, and the many other personality and environmental factors that can directly reduce depression and other forms of psychiatric distress.

Further information

www.hprt-cambridge.org <http://www.hprt-cambridge.org> (Harvard Programme in Refugee Trauma)

http://mentalhealth.samhsa.gov/cmhs/SpecialPopulations/refugmhnew.asp (U.S. Substance Abuse and Mental Health Services Administration, Refugee Mental Health Division)

www.unhcr.org (United Nations High Commission on Refugees)

References

1. Office of the United Nations High Commissioner for Refugees. (1993). *The state of the world's refugees, 1993: the challenge of protection*, Vol. 9, p. 191. Penguin Books, New York.
2. Office of the United Nations High Commissioner for Refugees. (1997). *The state of the world's refugees, 1997–1998: a humanitarian agenda*, Vol. 12, p. 298. Oxford, England; Oxford University Press, New York.

3. Office of the United Nations High Commissioner for Refugees. (1992). *Handbook on procedures and criteria for determining refugee status: under the 1951 convention and the 1967 protocol relating to the status of refugees*, p. 93. Office of the United Nations High Commissioner for Refugees, Geneva.
4. Independent Commission on International Humanitarian Issues. (1986). *Refugees: the dynamics of displacement: a report for the independent commission on international humanitarian issues*, Vol. 18, p. 152. Zed Books, Atlantic Highlands, NJ, London.
5. Eitinger, L. (1961). Pathology of the concentration camp syndrome. Preliminary report. *Archives of General Psychiatry*, **5**, 371–9.
6. Eitinger, L. (1972). *Concentration camp survivors in Norway and Israel*, p. 199. Martinus Nijhoff, The Hague.
7. Eitinger, L. and Strom, A. (1973). *Mortality and morbidity after excessive stress*. Oslo University Press, Oslo, Norway.
8. Strøm, A.C.S. (1968). *Norwegian concentration camp survivors*. Universitetsforlaget; Humanities Press, Oslo, New York.
9. Thygesen, P. (1980). The concentration camp syndrome. *Danish Medical Bulletin*, **27**(5), 224–8.
10. Thygesen, P., Hermann, K., and Willanger, R. (1970). Concentration camp survivors in Denmark: persecution, disease, disability, compensation. A 23-year follow-up. A survey of the long-term effects of severe environmental stress. *Danish Medical Bulletin*, **17**(3), 65–108.
11. Levy, B.S. and Sidel, V.W. (1997). *War and public health*, Vol. 19, p. 412. Oxford University Press, New York.
12. Mollica, R.F. and Caspi-Yavin, Y. (1992). Overview: the assessment and diagnosis of torture events and symptoms. In *Torture and its consequences* (ed. M. Basoglu), pp. 253–74. Cambridge University Press, Cambridge.
13. Mollica, R.F., *et al.* (1998). The dose-effect relationships between torture and psychiatric symptoms in Vietnamese ex-political detainees and a comparison group. *The Journal of Nervous and Mental Disease*, **186**(9), 543–53.
14. Quiroga, J. and Jaranson, J. (2006). Politically-motivated torture and its survivors: a desk study review of the literature. *Torture*, **15**(2–3), 1–111.
15. Dohrenwend, B.P. (1998). *Adversity, stress, and psychopathology*, Vol. 15, p. 567. Oxford University Press, New York.
16. Susser, M. (1981). The epidemiology of life stress. *Psychological Medicine*, **11**(1), 1–8.
17. Swiss, S. and Giller, J.E. (1993). Rape as a crime of war. A medical perspective. *The Journal of the American Medical Association*, **270**(5), 612–15.
18. Research Triangle Institute. (2005). *Refugee health promotion and disease prevention toolkit*. SAMHSA Center for Mental Health Services, Editor. Rockville, MD.
19. Kinzie, J.D., *et al.* (1990). The prevalence of posttraumatic stress disorder and its clinical significance among Southeast Asian refugees. *The American Journal of Psychiatry*, **147**(7), 913–17.
20. Mollica, R.F., Wyshak, G., and Lavelle, J. (1987). The psychosocial impact of war trauma and torture on Southeast Asian refugees. *The American Journal of Psychiatry*, **144**(12), 1567–72.
21. HPRT. (1998). *Trauma and disability: long-term recovery of Bosnian refugees*. Harvard Program in Refugee Trauma, Cambridge, MA.
22. Mollica, R.F., *et al.* (1993). The effect of trauma and confinement on functional health and mental health status of Cambodians living in Thailand-Cambodia border camps. *The Journal of the American Medical Association*, **270**(5), 581–6.
23. Shrestha, N.M., *et al.* (1998). Impact of torture on refugees displaced within the developing world: symptomatology among Bhutanese refugees in Nepal. *The Journal of the American Medical Association*, **280**(5), 443–8.
24. Mollica, R., Sarajlic, N., Chernoff, M., *et al.* (2001). Longitudinal study of psychiatric symptoms, disability, mortality, and emigration among Bosnian refugees. *The Journal of the American Medical Association*, **286**(5), 546–4.
25. Marshall, G., Schell, T.L., Elliott, M.N., *et al.* (2005). Mental health of Cambodian refugees 2 decades after resettlement in the United States. *The Journal of the American Medical Association*, **294**(5), 571–9.
26. Lin, E., Ihle, L.J., and Tazuma, L. (1985). Depression among Vietnamese refugees in a primary care clinic. *The American Journal of Medicine*, **78**(1), 41–4.
27. Mollica, R., Caridad, K., and Massagli, M. (in press). Longitudinal study of posttraumatic stress disorder, depression, and changes in traumatic memories over time in Bosnian refugees. *The Journal of Nervous and Mental Disease*.
28. Brune, M., *et al.* (2003). Treatment of drug addiction in traumatised refugees. A case report. *European Addiction Research*, **9**(3), 144–6.
29. Westermeyer, J., Lyfoung, T., and Neider, J. (1989). An epidemic of opium dependence among Asian refugees in Minnesota: characteristics and causes. *British Journal of Addiction*, **84**(7), 785–9.
30. Westermeyer, J. (1995). Cultural aspects of substance abuse and alcoholism. Assessment and management. *The Psychiatric Clinics North America*, **18**(3), 589–605.
31. D'Amico, E., Schnell, T.L., Marshall, G.N., *et al.* (2007). Problem drinking among Cambodian refugees in the United States: how big of a problem is it? *Journal of Studies on Alcohol*, **68**(1), 11–17.
32. Goldfeld, A.E., *et al.* (1988). The physical and psychological sequelae of torture. Symptomatology and diagnosis. *The Journal of the American Medical Association*, **259**(18), 2725–9.
33. Westermeyer, J. (1981). Lao folk diagnosis for mental disorders: comparison with psychiatric diagnosis and assessment with psychiatric rating scales. *Medical Anthropology*, **5**, 425–43.
34. Bolton, P. (2001). Local perceptions of the mental health effects of the Rwandan genocide. *The Journal of Nervous and Mental Disease*, **189**(4), 243–8.
35. Simons, R.C. and Hughes, C.C. (1985). *The Culture-bound syndromes: folk illnesses of psychiatric and anthropological interest*. Culture, illness, and healing, Vol. 15, p. 516. D. Reidel; Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers, Boston, Hingham, MA.
36. Jablensky, A., *et al.* (1981). Characteristics of depressive patients contacting psychiatric services in four cultures. A report from the WHO collaborative study on the assessment of depressive disorders. *Acta Psychiatrica Scandinavica*, **63**(4), 367–83.
37. Ingstad, B. and Whyte, S.R. (1995). *Disability and culture*, Vol. 10, p. 307. University of California Press, Berkeley.
38. Ormel, J., *et al.* (1994). Common mental disorders and disability across cultures. Results from the WHO collaborative study on psychological problems in general health care. *The Journal of the American Medical Association*, **272**(22), 1741–8.
39. Mollica, R., Caspi-Yavin, Y., and Lavelle, J. (1996). The Harvard Trauma Questionnaire (HTQ) manual: Cambodian, Laotian, and Vietnamese versions. *Torture—Quarterly Journal on Rehabilitation of Torture Victims and Prevention of Torture*, (Suppl. 1), 19–42.
40. Mollica, R.F., *et al.* (1992). The Harvard Trauma Questionnaire. Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. *The Journal of Nervous and Mental Disease*, **180**(2), 111–6.
41. Willis, G.B. and Gonzalez, A. (1998). Methodological issues in the use of survey questionnaires to assess the health effects of torture. *The Journal of Nervous and Mental Disease*, **186**(5), 283–9.
42. Mollica, R.F., McDonald, L.S., Massagli, M.P., and Silove, D.M. (eds.) (2004). *Measuring trauma, measuring torture*. Harvard Program in Refugee Trauma, Cambridge, MA.
43. Mollica, R.F. and Lavelle, J. (1988). Southeast Asian refugees. In *Clinical guidelines in cross-cultural mental health* (eds. L. Comas-Diaz and E. Griffith), pp. 262–304. Wiley, New York.
44. Mollica, R.F. (2006). *Healing invisible wounds: paths to hope and recovery in a violent world*. Harcourt Press, New York.

45. Mollica, R.F., *et al.* (2004). Mental health in complex emergencies. *Lancet*, **364**(9450), 2058–67.
46. Hinton, D.E., *et al.* (2005). A randomized controlled trial of cognitive-behavior therapy for Cambodian refugees with treatment-resistant PTSD and panic attacks: a cross-over design. *Journal of Traumatic Stress*, **18**(6), 617–29.
47. Hinton, D.E., *et al.* (2004). CBT for Vietnamese refugees with treatment-resistant PTSD and panic attacks: a pilot study. *Journal of Traumatic Stress*, **17**(5), 429–33.
48. Basoglu, M., *et al.* (2004). Cognitive-behavioral treatment of tortured asylum seekers: a case study. *Journal of Anxiety Disorders*, **18**(3), 357–69.
49. Bolton, P., *et al.* (2003). Group interpersonal psychotherapy for depression in rural Uganda: a randomized controlled trial. *The Journal of the American Medical Association*, **289**(23), 3117–24.
50. Schnurr, P.P., *et al.* (2007). Cognitive behavioral therapy for posttraumatic stress disorder in women: a randomized controlled trial. *The Journal of the American Medical Association*, **297**(8), 820–30.
51. Lavelle, J., Tor, S., Mollica, R.F., *et al.* (eds.) (1996). *Harvard guide to Khmer mental health*. Harvard Program in Refugee Trauma, Cambridge.
52. Mollica, R.F., Tor, S. and Lavelle, J. (1998). *Pathway to healing*. Harvard Program in Refugee Trauma, Cambridge.
53. Heigel, J. (1994). Use of indigenous concepts and healers in the care of refugees: some experiences from the Thai border camps. In *Amidst peril and pain: the mental health and well-being of the world's refugees* (ed. A. Marcella). American Psychological Association, Washington, DC.
54. Mollica, R.F., *et al.* (2002). Science-based policy for psychosocial interventions in refugee camps: a Cambodian example. *The Journal of Nervous and Mental Disease*, **190**(3), 158–66.
55. Silove, D., Ekblad, S., and Mollica, R. (2000). The rights of the severely mentally ill in post-conflict societies. *Lancet*, **355**(9214), 1548–9.
56. UNICEF. (1990). *Children and development in the 1990s: a UNICEF source book on the occasion of the World Summit for Children*. UNICEF, New York.
57. Basoglu, M. (1992). *Torture and its consequences: current treatment approaches*, Vol. 13, p. 527. Cambridge University Press, Cambridge, New York.

7.10.2 Mental health services for homeless mentally ill people

Tom K. J. Craig

Definition and demography of homelessness and its link to mental illness

The term 'homeless' has been used to describe populations as diverse as those sleeping in the shelter of a cardboard box, to those sleeping on a friend's floor. Given such wide definition, it is not surprising that estimates of the numbers involved vary greatly from survey to survey and from one country to another. But regardless of the definition, there is consensus that the numbers of homeless people in most Western urban areas increased during the past two decades, reflecting a scarcity of low-cost housing, the erosion of traditional family networks, and downsizing in the organization and delivery of supportive services. Of all these factors, the shortage of affordable accommodation is the most important. For example, in England there has been a 40 per cent increase since 2002 in the number of households on waiting lists for social housing with estimates that a minimum of 20 000 housing units above current government targets are required to simply meet newly arising urgent need.[1]

Compared with a domiciled population, homeless people are less likely to have completed basic education, less likely to have ever held employment, and more likely to have experienced parental neglect and abuse in their childhood.[2]

Given the evidence linking homelessness to poverty and social disadvantage, it is hardly surprising that homeless people report higher rates of psychiatric disorder relative to the general population. While rates vary depending on the particular measure of mental illness adopted by each study and by the homeless population being investigated, most report major psychiatric disorder in 30 to 60 per cent of those using emergency shelters and sleeping rough. The prevalence of schizophrenia and other psychoses is particularly high amongst the middle-aged residents of long-stay hostels, while depression, generalized anxiety, and impulsive self-harm are more typically encountered in younger runaways and adolescent populations. Alcoholism and drug dependency are present in as many as two-thirds of men and a third of homeless women. Co-morbidity of mental illness and substance use disorder is the rule rather than the exception as are the co-occurrence of respiratory disease, infections, trauma, and the physical consequences of poor diet, poor hygiene, and the complications of substance abuse. Of growing concern is the accumulation of older multiply disabled populations in some North American cities.[3]

The typical pattern of service utilization of the severely mentally ill among the homeless population is one of extremes—bursts of involuntary hospital admissions and compulsory treatment interspersed with long periods of neglect and isolation. Many of those who are found sleeping rough or resident in temporary shelters have found their way to these locations as a conscious effort to avoid contact with health and social care professionals and remain unwilling to be part of any structured rehabilitation programme.

Barriers to care

Poverty and isolation

Very few homeless mentally ill people have satisfactory links to family or other supportive social groups. Unemployment is the norm and many have histories of contact with the criminal justice system. The lack of supportive kinship networks mean that there is seldom anyone who has an interest in their welfare and no one on whom services can rely for informal care giving. Affordable housing is likely to be of poor quality and unsupervised. Landlords are reluctant to rent property to someone with a history of destructive behaviour, a criminal record, or manifest mental illness.

Barriers arising from the illness

Severe mental illness contributes to incompetence in many aspects of daily life, with impaired social function and problems initiating and executing daily living tasks that require a degree of forward planning. Co-morbid cognitive impairment or substance abuse compound these problems. Many homeless patients will have lost their accommodation as a direct result of their illness, being evicted

for failing to keep up with rent payments, neglecting or damaging the property, or following complaints from neighbours.

Barriers put up by services

The lack of common purpose and co-ordination of social welfare, health, and criminal justice agencies lies at the heart of many difficulties faced by homeless mentally ill people. A young homeless person, for example, may be too chaotic to undertake the retraining programme that is his only route to welfare support, may be unable to register with a local family health centre because of his lack of a permanent address, and may be summarily rehoused without reference to health services involved in his care. Finally, the prejudices of professionals can make services unacceptable and the emphasis on treatment is seldom attractive to a homeless person whose immediate needs are for food, shelter, and security.

Principles of service organization and delivery

To state the obvious, the solution to problems of homelessness lies in the provision of suitable accommodation, targeted efforts to re-house the most vulnerable, and sufficient longer-term tenancy support to prevent a return to the streets. While many countries have social welfare legislation to assist homeless people, only a minority provide a legally enforceable right to suitable accommodation for vulnerable populations. The Rough Sleepers Initiative in England, provides emergency accommodation for the roofless population and follows this emergency re-housing with a Tenancy Sustainment Programme of flexible practical and emotional support to prevent future accommodation breakdown.[4] This has been very successful in reducing the numbers of rough sleepers though gaps remain, particularly for people suffering from severe mental illness and substance dependency. For these populations, a further tier of service involving specialized mental health provision is needed either as part of an intensive initial stabilization[5] or in the longer-term. While the detail of specialized services varies according to local circumstances, they are all based on a small number of ideological and organizational principles (Table 7.10.2.1).

Improved inter-agency co-operation

As a first step, most involve a steering group comprising senior representatives of the key stakeholders in health, housing, social services, police, and voluntary sectors. These groups oversee the development of services across a wide geographical area—a large sector of a city or state. The members carry sufficient political and managerial authority to be effective in dealing with bureaucratic obstacles that are bound to arise from time to time.

Table 7.10.2.1 Services for homeless mentally ill people

Essential components for rehousing
Availability of temporary accommodation with a pathway to permanent housing
Capacity to deliver basic needs (shelter, food, income support)
Ongoing practical support to maintain tenancy
Specialized mental health service
Steered by a partnership of key stakeholders (housing, health, welfare, etc.)
Multidisciplinary front-line team
Assertive outreach model
Capable of managing mental illness and substance-use disorder
Wider context
Community-orientated mainstream psychiatric services
Influence of central and local government policies on health, welfare, and criminal justice

Providing local multidisciplinary specialist teams

This co-ordination is replicated at the local level through multidisciplinary clinical teams, joint working, and case management. Such partnership ventures have been established in several cities in North America, Europe, and Australia using a variety of organizational approaches ranging from a single multidisciplinary team through 'one-stop shops' where professionals from a variety of backgrounds come together to provide services at a common location.

Essential components of the specialist service

The management of a homeless mentally ill person involves stages of engagement, stabilization, resettlement, and the eventual transfer of care to mainstream providers. Engagement can take a long time, staff must be prepared to leave the clinic and go to where homeless people congregate. Help with welfare and practical problems may be all that can be done at first but the duty to maintain a therapeutic focus must always be maintained. Stabilization requires the specialist assessment and treatments provided to any mentally ill person, including hospital admission if necessary. The task of resettlement typically involves a compromise between personal preferences, available resources, and the level of support needed to promote rehabilitation. For example, independent accommodation may be a person's first choice but may only be a viable prospect if it can be backed up by weekly visits from the mental health team. Core and cluster arrangements, in which residents have their own flat but receive supervision from an on-site warden within the complex is a particularly effective model for those who have failed in independent accommodation but who reject shared facilities.

The eventual transfer of care to mainstream services can be quite difficult to manage and most follow-up studies suggest that fewer than half of those transferred remain in treatment.

Conclusion

Specialist multidisciplinary teams for homeless mentally ill people provide an essential safety net for those who have fallen out of the wider mental health care system. They offer distinct advantages in terms of their capacity to work across traditional geographical and bureaucratic barriers, to take the longer-term view of the task of engagement, and to bring together the multiple strands of care across different provider agencies. Introduced as a temporary measure over a decade ago, they are still with us and likely to remain a permanent fixture of urban mental health care.

Further information

Access to mental health services for people who are homeless or living in temporary or insecure accommodation. A good practice guide: http://www.communities.gov.uk/index.asp?id=1162512

Essential statistics on homelessness in Britain: http:// www.homeless.org.uk/policyandinfo/facts/statistics

References

1. National Housing Federation. (2006). *England's housing time-bomb: affordability and supply 2006–11*. National Housing Federation, London. Website: http://www.housing.org.uk/uploads/file/campaigns/tb_england.pdf
2. Bhugra, D. (ed.) (1996). *Homelessness and mental health*. Cambridge University Press, Cambridge.
3. Hahn, J.A., Kushel, M.B., Bangsberg, D.R., *et al.* (2006). The aging of the homeless population: fourteen-year trends in San Francisco. *Journal of General Internal Medicine*, **21**, 775–8.
4. Lomax, D. and Netto, G. (2007). Evaluation of tenancy sustainment teams. Department of communities and local government. http://www.odpm.gov.uk/index.asp?d=1505917
5. Susser, E., Valencia, E., Conover, S., *et al.* (1997). Preventing recurrent homelessness among mentally ill men: a 'critical time' intervention after discharge from a shelter. *American Journal of Public Health*, **87**, 256–62.

7.10.3 Mental health services for ethnic minorities

Tom K. J. Craig and Dinesh Bhugra

Ethnicity, culture, and health care need

Services aimed at minority ethnic populations are all too often developed on the basis of conspicuous morbidity than on any real understanding of the diversity of ethnic minority communities and their wider health needs. For example, in England, while much has been written about ethnicity and psychiatric morbidity, the literature remains largely focused on African-Caribbeans and Asians, while the needs of the Irish, who comprise the largest ethnic community by migration in many parts of the UK are seldom explicitly addressed despite evidence of high rates of suicide and unexplained death many times in excess of the indigenous population.[1] In addition, the large numbers of asylum seekers and refugees who move around the world, brings an increased need for culturally sensitive services. But very few models exist for developing these. The principles of good practice indicate that the start has to be a clear knowledge of the population that will be accessing services and an appreciation of the complicating factors of social disadvantage, material deprivation, and poverty.

There is no doubt that social disadvantage and racial prejudice whether real or perceived are pivotal in determining not only the mental health of minority populations but also the pathways individuals and their families use in seeking help for ill health. Delays in help-seeking can also be due to the stigma of mental illness and to sufferers' fears that they will be misunderstood and mistreated because of differences in culture, language, and racist attitudes within the services. These factors may be more apparent in older individuals and those who were born outside the country who may not be aware of various options available to them. Studies over the past 30 years or more in Britain, the Netherlands, the United States, Canada, and Australia have shown that minority groups have lower access to mental health services, are less likely to receive care, and when they do this is more likely to be of a lower quality. Black people in the UK and the United States are more likely than white people to be compulsorily detained in hospital, to be screened for drug abuse, to receive higher doses of medication and physical rather than psychological therapies. They are over-represented compared with their numbers in the general population, whether in general wards, locked wards, secure units, court diversion schemes, special hospitals, or prisons.[2]

Some of these problems can be attributed to a lack of understanding on the part of mental health practitioners of the cultural beliefs, values, and practices of minority groups with consequent shortcomings in assessment, diagnosis, and the provision of care. While language can be a major obstacle, for many people from minority ethnic groups who speak English the problem is of communication rather than language. The power dynamic that is always present in any clinical consultation is magnified and both patient and doctor will have predetermined expectations of how their interaction will turn out depending upon their experience of previous consultations. Problems in the interaction are likely to be interpreted as arrogance and racism on the one hand and indifference, wariness, or docility on the other. Thus, both missed diagnoses and misdiagnosis may result. A lack of recognition of the personal, social, and cultural problems which influence the presenting patterns of symptoms in different ethnic groups can contribute to the tendency of clinicians to make assumptions and listen out for stereotypical triggers which then prompt a particular therapeutic response. Such triggers include religious euphoria, use of Cannabis in African-Caribbeans, and the 'fatalistic attitude' attributed to Asian patients.

Culturally competent services

The past decade has seen an emerging consensus that the way forward lies in the development of fully integrated multicultural services with good working links with the local minority community rather than separate services. Such services would provide staff who can understand their client's cultural background and the ways in which this influences the presentation of distress and disorder. There would be closer working links with religious leaders and healers of local ethnic minority communities, female-only areas on wards, and a greater involvement and support of the family in understanding the problems and developing solutions.

In North America,[3,4] Britain,[5] and Australia[6] there has been a significant 'top–down' pressure to shift health care organizations in this direction and several large-scale programmes such as the European 9-country 'Migrant-Friendly Hospitals' initiative[7] have been reported. Although differing in detail, all these programmes share common elements. These principles are outlined in Table 7.10.3.1 setting out the main conclusions in a 'bottom–up' approach in order to emphasize the importance of changes in the attitudes, knowledge, and skills of front-line managerial and clinical staff.

Commonly referred to as 'cultural competence' these attributes include attention to obvious language differences but go further to include history, traditions, beliefs, and values even if the latter differ from those held by the professional. A culturally competent clinician is sensitive to a patient's cultural influences, expressions of distress and help-seeking, and is also aware of their own attitudes and prejudices and how these are in turn shaped by their own cultural background. This objective is achieved through

Table 7.10.3.1 Organizational steps towards a culturally competent mental health service

1 **Workforce level**
Training in 'cultural competence' should be mandatory for all mental health professionals
(a) Undergraduate programmes
(b) Post-graduate as continuing professional development
2 **Health care provider level**
Provide accommodation, washing, and living space facilities that take into account different cultural and gender definitions of ordinary social behaviour, dignity, and respect.
Senior management responsibility and accountability for:
(a) Active race equality policies
(b) Recruitment policy—to increase presence of minority staff and provision of training and support as needed
(c) Ensure staff have received relevant cultural competency training
Ensure adequate data collection includes a robust estimate of the numbers of ethnic minorities using services with a focus on key areas such as disparities by ethnicity in the use of coercive treatments, dropout from follow-up; differences in the uptake of psychological and pharmacological treatments.
Specialized outreach services targeting mentally ill people in the criminal justice system, homeless, and refugee populations.
Partnership arrangements with NGOs including the provision of volunteer advisors and of translation services that do not rely on relatives or other informal carers.
3 **Wider health service level**
At the appropriate Regional or National HMO or Statutory Organization:
(a) Policy and practice commitment to removing barriers to access, e.g. extend health insurance to the uninsured and closer integration of primary and secondary mental health care
(b) Collection of good quality demographic information and ethnic monitoring for planning and overseeing services
(c) Minority representation at all levels of health service planning and delivery
Continue to expand the science base to determine what works best for whom

training and a plethora of different cultural competency/diversity courses have sprung up in recent years. A search of the Internet identifies courses in undergraduate nursing, medical, and pharmacy programmes, in post-graduate continuing professional development and as part of wider organizational change and development. While the moral argument for improved cultural competence is hardly contestable, whether or not these training courses are sufficient to effect lasting change in behaviour and the delivery of health care is less certain. The very limited empirical evidence base is predominantly from the United States and shows efficacy in terms of short-term changes in attitudes and knowledge. Evidence is still lacking on which elements of this training are essential and on the downstream effects on service quality that is the real target.

Table 7.10.3.1 also shows the other key steps that have been taken towards developing more culturally sensitive services. Most health service providers have gone some way to providing female-only inpatient wards, more community-orientated service settings and addressing the need for expanding the numbers of people from minority backgrounds in their workforce though it is still all too easy to think that problems with cultural sensitivity can be solved with this alone. Simply hiring people on the basis of nationality, ethnicity, or skin colour will not necessarily ameliorate problems in the service. Typically these staff are the lowest in the hierarchy of power and have the least capacity for influencing either the care of the patient or wider attitudes within the institution. Even when employed in a position of power there is a danger of their appointment being seen as tokenistic.

Another important step involves the elaboration of performance indicators to assess the impact of training and service changes and ethnic monitoring with a focus on key areas such as the use of coercive treatments, treatment discontinuation, and ethnic differences in the uptake of psychological and pharmacological treatments. In the most sophisticated systems this includes both quantitative epidemiological and economic data as well as qualitative inputs from consultation with users, carers, and the general public including community and religious leaders.[8]

Finally, at the wider community level, NGOs have often become the champion of good practice, seen by their users as an antidote to inadequate mainstream care. These small organizations are generally based on consultation with users, carers, and local mental health professionals and are more in tune with the expressed needs of the community. The best have good working relations with mainstream services, and are generally seen as making an important contribution to wider community care. They provide a range of supportive services and are ideal vehicles for health promotion and dissemination of health-related information. Given the need to develop ways of working which promote inclusion of patients, their families, and the community in general, it is impossible to overstate the importance of effective liaison between the voluntary sector and mainstream psychiatric services. People often come into contact with the emergency services because there is a lack of knowledge of the availability of community alternatives and of where to go when distressed. Working with voluntary services will not only contribute to their longevity but also ensure that the complementary treatment modalities they offer are part of the service provided.

Even where these system-wide initiatives are absent or not yet fully implemented, individual clinicians can do a lot to improve the care of minorities within their own service. Finding out the scale of the problem is a good place to start. Are population estimates available from a recent census, is information available from a recent census or from local government sources? More importantly, are there known problems of access for these populations in general or particularly within local mental health services? Are significant numbers not treated because they are held in 'inappropriate' settings in the criminal justice system or because they are mobile populations? In terms of delivering the service, outreach is likely to be the key and this will almost always involve partnership arrangements, developing understanding of key cultural influences including the importance of alternative health perspectives, spirituality and traditional healing. Several innovative services have been developed alongside organizations that 'hold' significant numbers of the minority population such as the Church or the local Mosque or NGOs dealing with specific communities, which can provide a wider social or housing service to minority populations. It is the near universal experience of these services that initial progress is slow with many barriers of mutual misunderstanding and suspicion to overcome. Developing a relationship or even employing someone from the minority culture as a go-between, advisor, or 'cultural consultant' can be helpful though it needs careful preparation to avoid selecting a consultant from the wrong tribal

background or at the wrong level of seniority, gender, or language group. Culture broker models have been used in the United States, whereas cultural consultation services happen in parts of the UK and cultural liaison officers are used in parts of Australia. Whatever the choice, it is likely to take time and patience to develop a high level of 'visibility' in the target community as well as a sound understanding of its culture, taboos, and historical context. Where available, cultural supervision from a more experienced practitioner is also helpful. The underlying principle has to be a two-way process, which deals with information from the patient's community and the service provider to ensure that communications are clear.

A few further points can usefully be taken into account in planning the service. First an understanding of local models of illness that determine when and how and of whom help will be sought, second use of local epidemiological data to assess the impact of age and gender on service demand and finally the involvement of the local community to promote a sense of ownership and involvement in the delivery of services.

In conclusion, the diagnosis and treatment of mental ill health among multiethnic populations is probably one of the most complex and contentious challenges in psychiatric service provision. At the heart of this complexity are problems of ignorance, attitude, and failures of communication on all sides. It is also potentially one of the most rewarding endeavours if got right. No one service model is likely to apply to every community, even if people belong to the same ethnic group. The development of specialist psychiatric services may not always be possible or even essential. Instead, the requirements are for approaches that are flexible, sensitive, accessible, and accountable to the people they serve.

Further information

American Psychological Association guidelines for providers of services to ethnic linguistic and culturally diverse populations: http://www.apa.org/pi/guide.html

Sainsbury Centre for Mental Health. (2002). Breaking the circles of fear http://www.scmh.org.uk

References

1. Bhugra, D. (2004). *Culture and self-harm*. Psychology Press, Hove.
2. Lipsedge, M. (1993). Mental health: access to care for black and ethnic minority people. In *Access to health care for people from black and ethnic minorities* (eds. A. Hopkins and V. Bahl), pp. 169–83. Royal College of Physicians, London.
3. Office of Minority Health (USA). Assuring cultural competence in health care: recommendations for national standards and an outcome-focused research agenda. www.omhrc.gov/Assets/pdf/checked/Assuring_Cultural_Competence_in_Health_Care-1999.pdf
4. Mental Health: Culture Race and Ethnicity. A supplement to Mental Health: A Report of the Surgeon General. www.surgeongeneral.gov/library/mentalhealth/cre/
5. Department of Health. (2005). *Delivering race equality in mental health care: an action plan for reform inside and outside services and the government's response to the independent inquiry into the death of David Bennett*. HMSO, London. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4100773
6. Royal Australasian College of Physicians. (2005). Policy statement: aboriginal and Torres Strait Islander Health. Available at: http://www.racp.edu.au/index.cfm?objectId=49F4E2A9-2A57-5487-D0597D1ED8218B61
7. Migrant-Friendly Hospitals Initiative. (2004). Available at: http://www.mfh-eu.net/public/home.htm
8. Jordan, J., Dowswell, T., Harrison, S., *et al*. (1998). Health needs assessment. Whose priorities? Listening to users and the public. *British Medical Journal*, **316**, 1668–70.

SECTION 8

The Psychiatry of Old Age

This page intentionally left blank

8.1

The biology of ageing

Alan H. Bittles

Introduction

Although old age is readily recognizable, methods to define and measure the underlying biological processes are much less amenable to study. For this reason, **life expectancy** has been widely used as a surrogate measure of ageing, as well as to monitor economic progress at national and regional levels. It is generally acknowledged that lifespan is a constitutional feature of the human phenotype, and twin studies have indicated that 25–33 per cent of the variance in human **longevity** is genetic in origin.[1,2] External factors including lifestyle can also exert a major influence, as illustrated by the current mean life expectancies of 79 and 86 years for males and females in Japan, whereas the comparable figures for Botswana are 35 and 33 years, respectively.

The importance of genetic inheritance as a determinant of extended survival has been illustrated by population level studies in Okinawa, an island prefecture of southern Japan with a very high prevalence of long-lived individuals. On the island, the mortality rates of the male and female siblings of centenarians were approximately half those of birth cohort-matched, non-centenarian siblings.[3] These findings parallel an earlier study of the family of Jeanne Calment, who died in France in 1997 aged 122 years. Of her 55 relatives, 24 per cent had lived to >80 years compared to just 2 per cent of a matched control group.[4] However, it remains unclear whether the enhanced lifespan of individuals who exhibit above average longevity is due to a slowing of the overall ageing process or is primarily associated with resistance to major life-threatening pathologies.

The concept of an '**allostatic load**', potentially involving the neuroendocrine, sympathetic nervous, immune and cardiovascular systems, and metabolic pathways, has been advanced to describe the lifetime costs of adapting to physical and psychological stresses. According to this hypothesis, while the actions of biological mediators of **stress** can be initially beneficial to health, chronic stimulation results in regulatory imbalance and subsequent pathophysiological changes.[5] Empirical studies have indicated increased physiological dysregulation and functional decline at >70 years of age, which would imply that predicted global increases in the numbers of older persons will be accompanied by disproportionately larger groups of individuals with major age-related pathologies.

Theories of ageing

While initially popular, it became apparent that single, 'magic bullet' causes of ageing were inappropriate to complex biological species, and with this recognition earlier organ- and system-based theories have gradually been discounted. Conversely, the observation that ageing appears to be initiated at different ages and can proceed at different rates in individual members of a species provides presumptive evidence for the interaction of multiple genetic and non-genetic influences. Two main groups of theories have been formulated, genomic and stochastic, each subdivided into a number of discrete topic headings.

Genomic theories of ageing

Genomic theories premise that ageing is primarily associated with changes in the genetic constitution of the organism. Support for genomic theories stems largely from the characteristic life expectancies of mammalian and non-mammalian species, and theories proposing a primarily genetic basis for ageing were greatly strengthened by the demonstration that human diploid cells exhibited a highly reproducible lifespan when cultured in the laboratory. Although strong evolutionary advantages can be envisaged for genetic control of developmental changes up to and including reproductive adulthood, the existence of genes uniquely encoding ageing seems improbable since few free-living animals or humans have ever succeeded in attaining the maximum lifespan of their species.

(a) Information transfer

The ability to synthesize functional proteins is dependent on the fidelity of genetic information encoded in the DNA, its unimpaired transcription from DNA to RNA, and translation into peptides and proteins. As each of these processes is subject to inaccuracy, and during the life course of an organism the sequence of information transfer steps is continuously operational, the error potential is large. With increasing chronological age the probability of errors increases, resulting in the accumulation of deleterious mutations late in life.[6] Since a number of the proteins synthesized may be involved as surveillance enzymes to maintain the accuracy of the entire system, feedback mechanisms could lead to its collapse, resulting in a phenomenon initially termed **error catastrophe**.

(b) Somatic mutation

With the demonstration of an inverse correlation between the lifespan of mammalian species and the incidence of chromosome abnormalities, age-related physiological changes were originally ascribed to accumulated mutations in the nuclear DNA (**nDNA**) of somatic cells. Findings of this nature could, however, be explicable in terms of the ability of an organism to tolerate DNA damage via the repair of damaged molecules, with more than 130 human **DNA repair genes** identified.[7] The capacity of nDNA to resist attack by endogenous reactive species and environmental agents is therefore considerable, which casts doubt on the general applicability of the theory.

(c) Epigenetic mechanisms

Epigenetic errors, i.e. errors in the control of gene expression rather than mutations in DNA or protein, have been proposed as major primary causal factors in senescence.[8] In promoter regions of genes, hypermethylation silences a gene whereas the hypomethylation of previously methylated sequences permits their expression. The pattern of **DNA methylation** is established during development and is cell type-specific, and changes in methylation can occur both during ageing and in cancer cells. The advantage of epigenetic models of ageing is their lack of requirement for the evolutionary preservation of genes encoding ageing, which in former generations would seldom have been expressed.

(d) Mitochondrial decline

Mitochondria are subcellular organelles responsible for aerobic energy production in humans and many other species. The mitochondrial genome of ~16.5 kb is characterized by its extremely compact organization, with no protective histones, a lack of excision or recombinational repair mechanisms, and a virtual absence of introns, all of which make it highly susceptible to mutation. Mitochondrial DNA (**mtDNA**) plays a central role in mitochondrial propagation and the maintenance of cellular respiration, but a majority of proteins involved in the regulation of mtDNA transcription, translation and replication, and the mitochondrial respiratory chain, are encoded in the nuclear genome. This design requires the operation of a highly coordinated mechanism for the expression of the nuclear and mitochondrial genomes. The central role of mitochondria in energy production means that defects may be of major metabolic significance, and the demonstration of increased levels of mtDNA deletions and base-substitutions in aged human neurones, heart, and skeletal muscle suggest a causative role for mtDNA mutations in ageing.[9]

(e) Telomere loss

Telomeres are specialized structures located at the terminus of the DNA helix and critical to the maintenance of DNA stability and replication. The enzyme telomerase which is responsible for telomere synthesis is active during early embryonic and foetal development but its activity is down-regulated in all human somatic cells before birth. As human diploid fibroblasts in culture were shown to progressively lose telomeres, it was hypothesized that telomere length could act as a predictor of the potential *in vitro* lifespan achievable by a cell strain.[10] Humans have a common telomere profile found on lymphocytes, amniocytes, and fibroblasts which appears to be preserved throughout life. However, the rate of telomere loss with ageing varies between chromosomes and there is evidence that, in addition to the common human telomere profile, each person exhibits an individual profile.[11]

Besides ageing, telomere loss has been implicated in a wide range of disease states, including heart disease, stroke, infection, long-term chronic stress, and obesity. Given the apparent relationship between telomere loss and both ageing and age-related pathologies, pharmacological activation of telomerase has been proposed as a potential treatment for chronic or degenerative diseases.[12] As tumour tissue and transformed cells constitutively produce telomerase, any therapeutic intervention of this nature would require careful monitoring.

Stochastic theories of ageing

Stochastic theories of ageing propose that cumulative adverse random changes at the cellular level ultimately overwhelm the capacity of an organism to survive, with ageing representing the preceding period of functional decline.

(a) Rate of living

An optimum lifespan was achieved by a variety of non-mammalian species when the organisms were maintained at suboptimal temperatures. The further demonstration of an inverse relationship between basal metabolic rate and longevity in mammals was interpreted as evidence that a species lifespan was governed by its **rate of living**, which in turn was correlated with its level of energy expenditure. Theories of this type tend to be imprecise in defining the nature of the factor(s) controlling ageing and lifespan, although it was subsequently proposed that the rate of living theory could be reformulated as a **stress** theory of ageing, with stress resistance and longevity positively correlated.

(b) Waste product accumulation

Ageing has been ascribed to interference by accumulated waste products in normal cellular metabolism and function, ultimately resulting in dysfunction and death at cellular and organ levels. Lipofuscin, a highly insoluble, pigmented compound derived by auto-oxidation from incompletely degraded cellular materials and detected with advancing age in neurones, cardiac muscle fibres, and the adrenal cortex, has been particularly implicated. Alternatively, the build-up of lipofuscin in older organisms may be secondary to an age-related decline in the function of cellular catabolic processes.

(c) Macromolecule cross-linkage

Many macromolecules of biological importance develop cross-links with increasing chronological age. The establishment of cross-linkage, whether covalent in nature or due to hydrogen bonding, alters the chemical and physical properties of molecules. Thus cross-linkage of the extracellular protein collagen is believed to be responsible for the loss of elasticity in mammalian blood vessels and skin with advancing age, even though collagen is subject to turnover throughout the lifespan. DNA and RNA also are believed to be potential intracellular targets for cross-linking agents, and changes in their structure could have serious functional implications for cellular information flow.

(d) Post-synthetic modification

In addition to cross-linkage, molecular aggregation and immobilization that compromises cellular metabolism and function could be caused by post-synthetic modification of proteins, with non-enzymic glycosylation (glycation) particularly associated with ageing. Glycation is initiated by the reaction of glucose with the amino group of lysine residues, which then proceed to form a Schiff

base, and progressively more complex compounds collectively termed advanced glycosylation end (AGE) products. As little variation was found in the glycation levels of lens crystallin proteins in subjects aged between 10 and 80 years, post-synthetic mechanisms may be as much an effect as a cause of ageing.

(e) Free radical damage

The role of **free radicals** in ageing was first proposed over 50 years ago.[13] A wide range of highly reactive free radicals are derived from molecular oxygen, including the superoxide and hydroperoxyl radicals, hydrogen peroxide, hydroxyl radical, and singlet oxygen. The polyunsaturated fatty acid side chains of cell and organelle membranes form highly susceptible targets for the action of **reactive oxygen species (ROS)**, and the resulting lipid peroxidation can result in severe membrane damage and eventual death of the cell. DNA may also be a critical target molecule for free radical damage, with mtDNA especially susceptible because of its proximity to the site of free radical production in the inner mitochondrial membrane.[9] Although a wide variety of antioxidants have been identified in humans, including ascorbate, α-tocopherol, β-carotene, glutathione, and the enzymes superoxide dismutase, peroxidase, and catalase, there has been little experimental evidence that these antioxidants can produce a significant extension in maximum lifespan. However, transgenic mice expressing the free radical scavenger enzyme catalase targeted to mitochondria showed an approximately 20 per cent extension in their mean and maximum lifespans, and concomitant delays in cardiac pathology and cataract development.[14]

Ageing as an energy crisis

From an evolutionary perspective, it was suggested that senescence was the end-result of an energy conservation strategy operating in somatic cells. During the course of a lifespan, total available energy has to be allocated to a variety of functions, including macromolecular synthesis and degradation, cell and organ maintenance, and reproduction of the species. Since the energy supply is finite, and to ensure propagation of the species by the successful transmission of genes to future generations, a compromise has to be reached between the energy made available for each of these functions. According to the **disposable soma theory**, this accommodation in energy saving is achieved by maintenance of absolute or near absolute accuracy in germ cell replication but less rigorous error correction in somatic cells.[15]

As an organism ages, the demands placed on the free energy pool alter and increase from a primarily anabolic role to meeting the requirements of ever-increasing repair and catabolic functions, including those imposed by specific disease-related insults. If the mitochondrial inner membrane and/or mtDNA is damaged, an organism must increasingly rely on alternative, less efficient pathways for its energy needs, ultimately resulting in a critical shortfall in the energy supply needed to sustain life. In such an **energy crisis**, the somatic cells primarily affected would be post-mitotic cells with high energetic demands, typified by the heart, skeletal muscle, and the brain.

Dietary modification of ageing

Inherited factors clearly play a major role in ageing, and in determining the human lifespan. But if ageing also is stochastic in nature then it should be possible to modify development of the ageing phenotype by altering the relative influence of contributory environmental variables, including diet.

Dietary (or calorie) restriction, based on a diet reduced in total amount but otherwise nutritionally adequate, is the only method so far proven to increase maximum lifespan in mammals. The original dietary restriction experiments conducted in the 1930s resulted in animals that remained prepubertal as a result of their retarded growth and development.[16,17] In more recent food restriction experiments, rodents have typically been fed a diet corresponding to approximately 60 per cent of the food ingested by *ad libitum* fed controls, commencing either soon after weaning or in young adulthood. Under these circumstances, besides a increase in maximum lifespan the development of tumours and other chronic diseases of late adulthood was slowed.

DNA microarray studies into the effects of **calorie restriction (CR)** in mice have indicated a shift in transcriptional patterns towards increased protein turnover and decreased macromolecular damage. Rats maintained on a restricted diet (representing 60 per cent of the control diet) for 36 weeks showed increased transcription of muscle genes involved in ROS scavenging, tissue development, and energy metabolism, with decreased expression of genes involved in signal transduction, stress response, and structural and contractile proteins.[18] CR also was able to maintain ATP production but at the same time reduce age-dependent endogenous oxidative damage.[19]

Preliminary studies conducted on rhesus macaque monkeys aged approximately 20 years and maintained on a reduced caloric intake for 9–10 years have suggested that long-term CR produced beneficial alterations in glycogen metabolism and mitigated the development of insulin resistance in older animals. Although restricted numbers of primates have been studied and few have attained extreme old ages, a wide range of potentially beneficial outcomes have been reported, in particular an improvement in glucose tolerance, a lower core body temperature, an attenuated decline in dehydroepiandrosterone (DHEA) sulphate levels, decreased triglycerides and increased HDL2b, in combination with lower weight, lean body mass and fat, and lower energy expenditure (reviewed in Bittles 2008).

A number of small-scale human studies have been reported, including a 6-day investigation on eight adults and eight pubertal children involving a 50 per cent caloric reduction that resulted in a significant reduction in the nitrogen balance of both adults and children and a decrease in their insulin-like growth factor-1 (IGF-1) levels. Individuals who had voluntarily adopted a restricted food intake for 6 years displayed a wide variety of physiological, metabolic, and biochemical changes, all of which would be consistent with protection against atherosclerosis,[20] and CR alone or with accompanying exercise regimes significantly increased the numbers of mitochondria in skeletal muscle cells while decreasing both energy expenditure and the frequency of mtDNA damage in overweight healthy adults.[21]

Ageing and the concept of healthy life expectancy

Active Life Expectancy (ALE) is defined as the period of life free of disabilities which interfere with basic Activities of Daily Living (ADL), e.g. eating, getting in and out of bed, bathing and toiletry needs, dressing, and indoor mobility. The concept of healthy life expectancy has been extended by weighting specific physical and

cognitive dysfunctions to measure Disability-Adjusted Life Years (DALY) and Quality-Adjusted Life Years (QALY). **Disability Adjusted Life Expectancy (DALE)** is now widely used in epidemiological studies to estimate the number of years that might be expected to be spent in 'full health'. A common finding in developed countries was that although females enjoyed higher DALE scores they also could expect more years of disability at advanced ages. Although measures such as DALE, DALY, and QALY have been criticized on methodological grounds, given global increases in the numbers of elderly individuals, the concept of healthy life expectancy may be increasingly useful in identifying the health and support needs of the aged.

Discussion

As in other areas of medical science, the Human Genome Project has impacted strongly on research into biological aspects of ageing, and DNA analysis now offers major insights into the development of the ageing phenotype. During the last decade, DNA microarray studies have been adopted to investigate changes in gene expression that accompany ageing.

Initial studies on rodent tissue showed differential gene expression patterns with advancing age, indicative of a marked stress response and lower expression of metabolic and biosynthetic genes. An 'ageing transcriptome' conserved across mammalian species has been identified, comprising deregulation of mitosis, cell adhesion, transport, signal transduction, mitochondrial function, and inflammatory response, and accompanied by a reduction in processes dependent on energy metabolism and mitochondrial function.[22] Subsequent analysis of human tissue based on large-scale DNA microarrays has revealed diverse patterns of both increased and decreased gene expression. However, a study of ~32 000 muscle tissue genes obtained from volunteers aged 16 to 89 years confirmed the existence of a common ageing signature, with altered levels of expression in 250 age-regulated genes and three genetic pathways that correlated both with chronological and physiological age.[23]

Research on adult stem cells may provide key future insights into ageing. Adult stem cells mainly undergo chronological ageing, as in skeletal muscle, or exhibit a combination of chronological and replicative ageing, typified by haematopoietic stem cells.[24] What remains to be determined is whether the overall decline in tissue regenerative capacity with advancing age is caused by intrinsic ageing of stem cells, or is due to increasing impairment of stem cell function in an aged tissue environment. Until this basic question is resolved, the prospect of stem cell therapy as a potential 'treatment' to correct the functional declines and degenerative diseases typical of human ageing will remain a theoretical possibility.

Further information

Agren, G. and Berennson, K. (eds.) (2007). *Healthy ageing—a challenge for Europe*. Swedish National Institute of Public Health, Stockholm; ISBN 91 7257 481 X.

http://www.healthyageing.nu.

Bittles, A.H. (2008). The biology of human ageing. In *Psychiatry in the elderly* (4th edn) (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas). Oxford University Press, Oxford, in press.

Johnson, M., Coleman, P.G., and Bengtson, V.L. (eds.) (2007). *Cambridge handbook of age and ageing*. Cambridge University Press, Cambridge; ISBN 052 182 6322.

National Institute of Aging (NIA); http://www.grc.nia.nih.gov/

Baltimore Longitudinal Study of Aging.

NIA Intramural Research Program 2006 Factbook.

References

1. Herskind, A.M., McGue, M., Holm, N.V., *et al.* (1996). The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics*, **97**, 319–23.
2. Ljungquist, B., Berg, S., Lanke, J., *et al.* (1998). The effect of genetic factors for longevity: a comparison of identical and fraternal twins in the Swedish twin registry. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, **53**, M441–6.
3. Willcox, B.J., Willcox, D.C., He, Q., *et al.* (2006). Siblings of Okinawan centenarians share lifelong mortality advantages. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, **61**, 345–54.
4. Robine, J.M. and Allard, M. (1998). The oldest old. *Science*, **279**, 1834–5.
5. Seeman, T.E., McEwen, B.S., Rowe, J.W., *et al.* (2001). Allostatic load as marker of cumulative biological risk: MacArthur studies of successful aging. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4770–5.
6. Hughes, K.A., Alipaz, J.A., Drnevich, J.M., *et al.* (2002). A test of evolutionary theories of aging. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 14286–91.
7. Wood, R.D., Mitchell, M., Sgouros, J., *et al.* (2001). Human DNA repair genes. *Science*, **291**, 1284–9.
8. Holliday, R. (1987). The inheritance of epigenetic defects. *Science*, **238**, 163–70.
9. Kujoth, G.C., Bradshaw, P.C., Haroon, S., *et al.* (2007). The role of mitochondrial DNA mutations in mammalian aging. *PLoS Genetics*, **3**, 0161–73.
10. Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* (*London*), **345**, 458–60.
11. Graakjaer, J., Londono-Vallejo, J.A., Christensen, K., *et al.* (2006). The pattern of chromosome-specific variations in telomere length in humans shows signs of heritability and is maintained through life. *Annals of the New York Academy of Sciences*, **1067**, 311–16.
12. Harley, C.B. (2005). Telomerase therapeutics for degenerative diseases. *Current Molecular Medicine*, **5**, 205–11.
13. Harman, D. (1992). Free radical theory of aging. *Mutation Research*, **275**, 257–66.
14. Schriner, S.E., Linford, N.J., Martin, G.M., *et al.* (2005). Extension of murine life span by overexpression of catalyse targeted to mitochondria. *Science*, **308**, 1909–11.
15. Kirkwood, T.B.L. (1977). Evolution of ageing. *Nature*, **270**, 301–4.
16. McCay, C.M., Crowell, M.F., and Maynard, L.A. (1935). The effect of retarded growth upon the length of lifespan and upon the ultimate body size. *The Journal of Nutrition*, **10**, 63–79.
17. McCay, C.M., Ellis, G.H., Barnes, L.L., *et al.* (1939). Clinical and pathological changes in aging and after retarded growth. *The Journal of Nutrition*, **18**, 15–25.
18. Sreekumar, R., Unnikrishnan, J., Fu, A., *et al.* (2002). Effects of caloric restriction on mitochondrial function and gene transcripts in rat muscle. *American Journal of Physiology. Endocrinology and Metabolism*, **283**, E38–43.
19. Lopez-Lluch, G., Hunt, N., Jones, B., *et al.* (2006). Calorie restriction induces mitochondrial biogenesis and bioenergetic efficiency. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1768–73.
20. Fontana, L., Meyer, T.E., Klein, S., *et al.* (2004). Long-term calorie restriction is highly effective in reducing the risk for atherosclerosis in

humans. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 6659–63.
21. Civitarese, A.E., Carling, S., Heilbronn, L.K., *et al.* (2007). Calorie restriction increases muscle mitochondrial biogenesis in healthy humans. *PLoS Medicine*, **4**, 0485–94.
22. Wennmalm, K., Wahlestedt, C., and Larsson, O. (2005). The expression signature of *in vitro* senescence resembles mouse but not human aging. *Genome Biology*, **6**, R109.
23. Zahn, J.M., Sonu, R., Vogel, H., *et al.* (2006). Transcriptional profiling of aging in human muscle reveals a common aging signature. *PLoS Genetics*, **2**, e115.
24. Rando, T.A. (2006). Stem cells, ageing and the quest for immortality. *Nature Genetics*, **441**, 1080–6.

8.2

Sociology of normal ageing

Sarah Harper

Introduction

Research on the sociology of normal ageing has focused on understanding the paradigms of '*successful ageing*'. In an apparent reaction to '*disengagement theory*'[1] which proposed that to withdraw from roles and relationships in old age was normal, a new conceptual framework was developed in the late 1960s and 1970s which attempted to explain how individuals adapted to the constraints of ageing and old age. This has been variously measured in terms of good health, high levels of physical and mental functioning, and active engagement with one's social and physical environment. While post-modernism and critical gerontology have attempted to refocus the debate, the emphasis of most research and writing has remained within the framework of understanding, explaining, and even facilitating, 'success' in old age.

There is also a body of research which recognizes the importance of the *life course perspective*, and that throughout an individual's life, he or she is faced with *continuities* and *discontinuities* which have to be negotiated and resolved. Old age is but part of this lifelong process. Changes which occur in later life, such as retirement and widowhood, will lead to discontinuities in roles and relationships, other aspects of our lives will undergo little change allowing continuity. Alongside this, perspectives from anthropology, history and the social constructionist school of thought have also been recently influential.

This chapter will discuss concepts of age, generation, and cohort. It will consider the contribution of the life course approach to understanding ageing, and the manner in which other perspectives, such as social constructionism, narrative psychology and anthropology, have contributed to the sociology of normal ageing.

Structuring the life course through age

According to Hazelrigg,[2] the concept of age introduces signposts which link memory and anticipation, an iteratively remembered past and an iteratively expected future. Age classification is thus integral to normal organization of consciousness. As Mead's extensive work on life history, reminiscence and autobiography informs us,

> *one interacts retrospectively with one's younger selves, recalling earlier states of selfhood in the productive functioning of memory, and interacts prospectively with ones' older selves, anticipating conditions, actions, goal realizations and the like of late states of selfhood.*[1]

For both the individual and society, age conveniently dissects the life course into more manageable components. As a capitalist, industrial system emerged, and individuals moved from domestic units to bureaucratically organized corporations, so age was used to define adulthood and thus labour force participation. Age became the basis for regulating a large population. It defined the responsibilities of citizenship, and for each age related transition there is a stage of preparation, a stage of participation, and a stage of retirement.

Various anthropological studies[2] have highlighted alternative ways in which the life course might be structured. One of the most influential anthropological studies on the sociology of ageing was Cowgill and Holmes[3] work on ageing and modernization, which argued that the marginalization of older people was directly linked to modernization. While extensively debated ever since, this work highlighted the importance and complexity of cultural diversity. The burgeoning of anthropological studies around the concept of age and ageing since the Cwgill and Holmes study have contributed significantly to our understanding of this diversity.

Neither the !Kung nor Herero, hunter-gather and Bantu pastoralist peoples respectively of Botswana, have a concept of chronological age, marking age by physical transitions. Alternatively, the Tuareg, a semi-nomadic peoples in northern Niger, noted age by social transitions—courtship, marriage, childbirth, and grandchildren. Here, life transitions defining the ageing process are predominantly social rather than biological. A girl becomes a women not at menstruation, but at marriage; a women becomes an older women not at menopause but on having a child marry. For the Sukama of north-west Tanzania, ageing is defined through life course events. This emphasizes the social status of elderhood, measured by the wealth of alliances, offspring and livestock, which could not be diminished through ill health or loss of mental capacity. The Gussui of south-western Kenya have a similar notion of elderhood. However, they have adapted this traditional seniority gradation based on networks and affiliations to modern demands,

---

1 G.H. Mead, quoted in Hazelrigg p. 105[2]

2 See Harper for a full reference list to these studies[5]

incorporating such aspects as the role of entrepreneur to the criteria for achieving successful seniority status.

Modern Japanese society still applies a wide variety of terms to different points of the life course indicating complex relationships between chronological age and life transitions and physical appearance. For example, mid-life men and women with children whether or not they are married, will commonly be referred to as uncle and aunt, (*oji-san* and *oba-san*). Similarly, old men and women are frequently given the name of grandfather or grandmother, (*ojii-san* and *obaa san*) regardless of the presence of grandchildren, a characteristic also found in some European countries such as Greece. It is therefore clear that the domination of chronological age, has less salience in some other cultures.

Generation and cohort

Two further important concepts are generation and cohort. Individuals born within the same time period may be perceived as having a shared history and a common biography. The concept of *generation* is thus the link between an individual life course and the social changes that occur during the historical time of that life course. A generation may thus be thought of as *embodied history*.

Many of these draw on ideas from Mannheim who explored the creation of society through the continuous emergence of new age groups or generations. He argued that if social processes were always carried on and developed by the same individuals then once established, any fundamental social pattern, attitude or intellectual trend would probably be perpetuated. Culture was thus developed by individuals who come into contact anew with the accumulated heritage, that is the role of generations and while the continuous emergence of new individuals results in some loss of accumulated possessions it facilitates re-evaluation of our inventory and teaches us both to forget that which is no longer useful and to covet that which has yet to be won.

The problem for quantitative social scientists is how to disentangle those factors pertaining to the individual life course from those emerging from the historical context. It is here that the concept of cohorts, and cohort analysis has been refined by some to form a more analytical tool in the understanding of age and generational change.

A cohort begins with a particular demography at birth, that is its sex, race and economic composition. Differential mortality may lead to a higher proportion of some sub-groups surviving to old age; social mobility may lead to changes in cohort social status composition; and different historical periods will allow or enhance differential migration in and out of specific cohorts. A more sophisticated analysis places cohorts within specific historical contexts.[4] The *life-stage principle* suggests that disruptive social changes have enduring consequences on the subsequent lives, a particularly marked effect on those vulnerable at the time of occurrence.

Life course perspective

The life course perspective views old age as part of a life-long process of continuity and change. These can be addressed within four main frameworks: *context, transitions, roles*, and *relationships*.

Context

A starting point for life course analysis is the acknowledgement of the historical *context* within which different cohorts experience different aspects of the life course to life course perspective. As Harper[5] explores, while, most older men experienced a long period of economic activity followed by abrupt retirement, many older Western women experienced their younger lives within a framework of primary domestic duties, supplemented by intermittent economic activity. As a result, most older women replaced low earning capacity or economic dependence in younger life, with low incomes in old age. Cohorts in mid life, however, have had very different social and economic frameworks within which to live out their lives. Half the labour force in many countries is now female and full-time economic employment, with or without domestic, in particular childcare responsibilities, is becoming a widespread experience for many women. Despite this, there are still considerable income disparities in earning capacity of mid-life men and women. However, it is likely that future cohorts of older women will have higher incomes relative to older cohorts, and a lower gender income disparity.

Transitions

The processes which occur within these contexts can be understood as a series of life transitions.[6] Key transitions associated with later life are the end of active parenting, grandparenthood, widowhood and retirement. Each of these phases of life which may overlap, may be understood in relation to prior phases, and are mediated by other variables such as gender, class, and race. The transition to grandparenthood, for example, is experienced very differently by men and women, while the end of active parenting and transition to parent of a non-dependent child, the so-called *empty nest syndrome*, is mediated both by gender and by the experience of active parenting itself.

The transition to *widowhood* is one of the most stressful events of later life, with a high prevalence of depression both immediately before (presumably due to anticipation of the event and/or associated care giving) and in the first year following bereavement. Widowhood is likely to lead to lower income but higher social contacts for women, while men maintain their income, but are more likely to lose social contacts, unless they remarry. Over half women over 65 are widowed, rising to four-fifths at 85. Only 17 per cent of men are widowed over 65, rising to 43 per cent by their late 80s. Nearly three-quarters of older men in the UK are married, compared with less than a half of older women. This is explained both by differential life expectancy and the tendency of men to remarry following divorce or widowhood.[7]

The transition to *Grandparenthood* is the current normal experience of old age. US data suggests that more than half the population aged over 55 are in four-generation families and three-quarters of this population are or can expect to be as grandparents, with a prediction that one-third of current grandparents will live to be great grandparents, and one-fifth of all women who reach 80 will spend some time in a five-generation family as great-great-grandmothers. A similar picture may be found in the UK with estimates that three-quarters of adults over 66 years of age are grandparents. The transition from parenthood to grandparenthood, and even great-grandparenthood, determines both an individual's self-identity and subsequent roles and functions as

grandparenthood. In addition, the experience of the relationship that the grandchild has with his or her grandparent earlier will partially determine the way they take on the role and relate to their own grandchildren later on in life. Other sociological theories have been applied to the study of grandparenthood. *Role theory* suggests that a successful transition to grandparenthood requires both some socialization to the role, and appropriate life course timing. *Social stress theory* is used to argue that stress associated with transition to grandparenthood is related to the number, type, and context of the transitions and moderated by gender, education, income, and race.[9]

For many individuals, especially men, the transition to *retirement* is abrupt. Although early retirement has increased in Europe over the past twenty five years, most men still retire from full time work in their early-to mid-60s. A successful transition to retirement requires securing both financial security and personal adjustment. Atchley[10] identifies several phases of retirement. Pre-retirement, which may include a combination of both negative and positive feelings towards the impending event; a honeymoon period immediately following the event, which may extend for several years depending on the adjustment and resources, social, financial, and personal, available to the individual; disenchantment and reorientation; and eventually (if successful) stability. This latter stage occurs with the development of a well established set of criteria for making choices and dealing with the challenges and opportunities of this new life phase.

Roles and relationships

The above transitions—retirement, widowhood, grandparenthood - are also, of course, phases of life with specific roles and relationships. Thus, the transition of retirement also has an associated phase of being retired; that of grandparenthood of being a grandparent; that of widowhood of being a widow or widower.

We can examine two of these—late-life parenting and grandparenting—in the context of negotiating transitions and continuities in family roles and relationships as individuals age.[11] Intergenerational solidarity—shared values, normative obligations and enduring ties—and intergenerational conflict—whereby issues are resolved and relationships move on have been long seen as important components of this. More recently the concept of intergenerational ambivalence has been introduced. This, it is argued, reflects the contradictions which occur with ageing within family relationships These arise both through the desire of parents and children for both help and freedom, and conflicting norms regarding family relationships especially around the issue of care giving.

(a) Late-life parenting

Increasing longevity also means that most parent-child relationships will be lived out as predominantly non-dependent adult dyads, this is despite the delaying of child birth. The common experience for many parents and children is around 60 years of joint life, of which under one-third is spent in the traditional parent/dependent-child relationship. Around one-quarter of UK women and nearly 40 per cent of US women aged 55–63 still have a surviving parent. These women have thus spent around 60 years a child, some 40 of them in an adult relationship with a living parent. This relies on *re-bonding* in adulthood sometimes also referred to as 'reverse bonding'.[12] Under such experiences we see a loosening of the association between marital and parental roles. As the common experience of parenthood moves to more than 50 years of shared life, parents and children are adjusting to spending most of their relationship as independent adults. Similarly, husbands and wives are spending fewer of their joint lives as parents of young children. Relationships which have been historically based on a hierarchy which existed in part to support successful reproduction must move to greater equality, both child-parent, and husband-wife, as traditional roles based on parenthood give way to companionate relationships.[12]

(b) Grandparenting roles[3]

Currently, women can expect to become grandmothers in their 50s and 60s due to the early first age of births in the 1960s and 1970s. In addition, grandparental roles are lasting far longer due to increased longevity, the grandparent is thus more likely to be able to build a relationship with their grandchild into their adulthood. As a result many grandmothers, in particular, now face simultaneous demands as children of frail and dependent parents, mothers and grandmothers, as well as still being in full or part time economic employment.[13]

Grandmothers, in particular maternal grandmothers, are repeatedly attributed with having more influence in almost every value domain over their grandchildren than grandfathers. Research into the role of grandfathers has been limited.[14] However, it has been proposed that men become more nurturing as they get older and it could be hypothesized that these qualities might be expressed in relationships with their grandchildren. Similarly, the need to consider grandfathers as important resources for teenage mothers who are rearing their children, has been stressed. Harper also found that grandfathers could act as replacement partners and replacement fathers in female single parent households.[13]

Various roles of grandparenthood have been identified.[15] Bengtson (1985) for example, identifies five separate symbolic functions of grandparents: being there; grandparents as national guard; family watchdog; arbiters who perform negotiations between (family) members; and participants in the social construction of family history. Harper's study of grandmothers identifies grandmother as carer, replacement partner (confidante, guide and facilitator), replacement parent (listener, teacher and disciplinarian), and as family anchor (transferring values, attitudes and history).[13]

Complementary perspectives

Our understanding of ageing from a life course perspective has drawn on valuable insights from narrative gerontology and the social constructionist perspective.

Narrative gerontology

Narrative gerontology's main contribution to the field of ageing has been to the role of the life story in the development of theoretical and empirical approaches to ageing. It presumes that individuals think and act on the basis of stories, which have an external structure and an internal reality. Individuals retell their stories as they progress through their lives to make sense of their lives, and from a sociological perspective this retelling when carried out publicly provides considerable insights into various aspects of an individual's experience of ageing across the life course. There are

[3] A full reference list on grandparenthood may be found in Harper.[5]

four dimensions to this: the structural story, which reflects the wider societal context inhabited across the life course; the socio-cultural story, which reflects other identities such as ethnicity and gender; the situational story which reflects an individual's roles and relationships; and the interpersonal story, the meanings which the individuals place themselves on their life story.

Writers in this genre of theory have further suggested that individuals experience two types of time. Achenbaum[16] describes this as a *physical outer time* and *a psychological inner time*, or as Kenyon and Randall[17] state *clock time* and *story time*. There may be a tension between the two types of time described as *on time* and *off time*.[18] Hazelrigg[19] suggests that the tension between the two times has been become more extreme within modern society, so dominated by rigid timekeeping. He argues that modern life is lived in two separate registers. On the one hand, most of a life experience is formed directly and indirectly in a highly standardized sequence of institutionalized events—schooling, work, parenting, retirement. These events are regulated by procedural rules and recognized routines, with predictable durations and regulated transitions between events. On the other hand, those aspects of life experiences that are not institutionalized and structurally stabilized in recognized life course sequences tend to have little or no connection to status dimensions or specific locations in the life course. These would include self-image, personal satisfaction, existential aesthetics etc.

Tensions arise when the two registers fail to coincide—*off time*.[18] Examples include middle-aged couples falling in love and publicly exhibiting displays of physical affection and romance, or older people adopting student style lives. Off time may also include the experience of being externally forced, through illness for example, to fall outside the normal behaviour range as defined for one's age. This also includes examples where society takes an individual and places them within a situation which is unusual for their chronological age–for example, a very young person being rapidly promoted within an institution which is very highly age regulated, such as a university, and taking on a professorial mantle.

Social constructionist approaches

The sociology of ageing has also recently enveloped social constructionist approaches, including phenomenology, symbolic interactionism, and ethnomethodology. These approaches share a subjective orientation to social reality, focusing on describing how individuals negotiate their worlds, rather than trying to explain why. A key area of interest here is the management of identity across the life course and in particular as we age. Examples include, Matthew's[20] seminal work using symbolic interactionist approaches to explore how old women negotiate their own identities when they are continually deluged with negative public stereotypes of infirmity and worthlessness. This is a theme taken up later by Featherstone and Hepworth[21] in their work on the Masks of Ageing. Similarly, Karp[22] uses symbolic interactionist approaches to explore the impact of social messages on the emerging consciousness of men and women in their 50s of their own ageing.

Conclusion

The sociology of, so-called, normal ageing thus combines perspectives from diverse traditions of thought. This actual measurement of ageing, in terms of health, physical and mental functioning, and active engagement with one's social and physical environment, has combined quantitative social science, epidemiology and political economy. The understanding of these interactions, draws on perspectives from sociology, psychology, and anthropology. In both cases, however, there is now recognition of the importance of context and process, and that, in reality, there is perhaps no single construct of 'normal' ageing. Given this perspective, it is essential that care for older adults is person-centred rather than imposed by professionals.

Further information

Johnson, M., Bengtson, V., Coleman, P. *et al.* (eds) (2005) *The Cambridge Handbook of Age and Ageing*. Cambridge University Press, New York.

Ageing Horizons: http://www.ageing.ox.ac.uk/ageinghorizons/index.htm

Binstock, R. and George, L. (2005). *Handbook of Ageing and the Social Sciences*. Elsevier, London.

References

1. Cumming, C., and Henry, W. (1961). *Growing Old*, New York Basic Books.
2. Hazelrigg, L. (1997). On the importance of Age. In *Studying aging and social change: conceptual and methodological issues* (ed. M. A. Hardy). pp. 93–128. Sage Publications, Thousand Oaks.
3. Cowgill, D. and Holmes, L. (eds.) (1972). *Ageing and Modernization*, New York, Apple Century Crofts.
4. Elder, G. H., and O'rand, A. M. (1995). Adult lives in a changing society. In *Sociological perspectives on social psychology* (eds. K. S.Cook, G. A. Fine, & J. S.House), pp. 452–475. Allyn and Bacon, Boston.
5. Harper, S. (2006). *Ageing Societies*. Hodder Arnold, London.
6. Elder, G. H. (1985). Perspectives on the life course. In *Life course dynamics: trajectories and transitions* (ed. G. H. Elder) *1968–1980*. pp.23–49. Cornell University Press, Ithaca.
7. Arber, S. Davidson, K. and Ginn, J. (eds.) (2003). *Gender and Ageing: changing Roles and Relationships*, Maidenhead, Open University Press.
8. Kornhaber, A. (1996). *Contemporary Grandparenting*, New York, Sage.
9. Szinovacz, M. (1997). Grandparents today: a demographic profile. *The Gerontologist,* **38**, 37–52.
10. Atchley, R. (1989). A continuity theory of normal aging, *The Gerontologist,* **29**, 183–90.
11. Giarrusso, R., Silverstein, M., Gans, D., *et al.* (2005). Ageing parents and adult children: new perspectives on intergenerational relationships. In *The Cambridge handbook of age and ageing* (eds. M.L. Johnson, and V. L. Bengston). Cambridge University Press, New York.
12. Harper, S. (2006). *Ageing Societies*. Hodder Arnold, London.
13. Harper, S. (2005). Understanding Grandparenthood. In *The Cambridge handbook of age and ageing* (eds. M.L. Johnson, and V. L. Bengston), Cambridge University Press, New York.
14. Mann, R. (2007). 'Out of the shadows?: Grandfatherhood, age and masculinities', *Journal of Aging Studies*, **21**(4), 271–81.
15. Bengtson, V. (1985) Diversity and symbolism in grandparental roles. In *Grandparenthood* (eds. V. Bengtson, and J. Robertson) Beverly Hills, Sage.
16. Achenbaum, W. A. (1991) Time is the messenger of the Gods: a gerontological metaphor. In *Metaphors of Aging in Science and Humanities* (eds. *G.* Kenyon, J. Birren, and J. Schroots), pp. 83–101. New York, Springer.

17. Keynon, G., and Randall, W., (1997). *Restorying our Lives: personal growth through autobiographical reflection*. Westport, CT, Praeger.
18. Hagestad, G. O. (1986). Dimension of time and the family. *American Behavioral Scientist*, **29**, 679–94.
19. Hazelrigg, L. (1997). On the importance of Age. In *Studying aging and social change: conceptual and methodological issues* (ed. M. A. Hardy), pp. 93–128. Sage Publications, Thousand Oaks.
20. Matthews, S. (1979). *The Social World of Older Women*. Newbury park, Sage.
21. Featherstone, M., and Hepworth, M., (eds.) (1995). *Images of Ageing*. Sage, London.
22. Karp, D. (1988). A decade of reminders: changing age consciousness between fifty and sixty years old, *The Gerontologist*, 727–38.

8.3

The ageing population and the epidemiology of mental disorders among the elderly

Scott Henderson and Laura Fratiglioni

In the last decades the ageing of the populations has become a worldwide phenomenon.[1] In 1990, 26 nations had more than 2 million elderly citizens aged 65 years and older, and the projections indicate that an additional 34 countries will join the list by 2030. In 2000, the number of old persons (65+ years) in the world was estimated to be 420 million and it was projected to be nearly 1 billion by 2030, with the proportion of old persons increasing from 7 to 12 per cent.[2] The largest increase in absolute numbers of old persons will occur in developing countries; it almost triples from 249 million in 2000 to an estimated 690 million in 2030. The developing regions' share of the worldwide ageing population will increase from 59 to 71 per cent. Developed countries, which have already seen a dramatic increase in people over 65 years of age, will experience a progressive ageing of the elderly population itself (see Fig. 8.3.1). The global trend in the phenomenon of population ageing has dramatic consequences for public health, health care financing, and delivery systems in the whole world. The absolute number of chronic diseases as well as psychiatric disorders is expected to increase. In this chapter, the epidemiological aspects of the most common psychiatric disorders of the elderly are summarized and discussed.

Depressive disorders

The epidemiology of depression in the elderly can be approached at three levels: its occurrence in the elderly living in the community, in those reaching primary care, and in the residents of hostels and nursing homes.

The community

It might be expected that, overall, the prevalence of depressive symptoms and disorders might increase in old age due to the loss of partners, friends, social status, retirement, income, and, above all, declining health. It is surprising, therefore, that surveys of the elderly in the general population have recurrently found rates that are significantly lower than in younger adults. Many of the large national surveys have not included persons aged over 65 years, but two exceptions are Australia, which found a 12-month prevalence of 1.7 per cent for the 65 years and over group compared with 5.8 per cent for all adults; and the New Zealand survey with rates of 2.0 and 8.0 per cent, respectively. It must be emphasized that these data refer to depressive symptoms in the elderly living in the community.

What is so far unproven is that such findings are indeed valid, and if they are, what might explain them.[3] They could be due to sample bias, in which elderly respondents with depressive symptoms may be more likely to decline to be interviewed than younger depressed people. Selective mortality has also been proposed, but cannot account for the size of the difference. It could be due to an error in case ascertainment, by which the interview instrument is not equally valid across age groups. For example, questions about depressive symptoms may be responded to differently by persons aged 20 and 80 years. Another possibility is a cohort effect in much of the Western world, where people born in the second half of the twentieth century have higher rates for depression.[4,5] This seems increasingly likely and may be due to a combination of social and environmental factors.

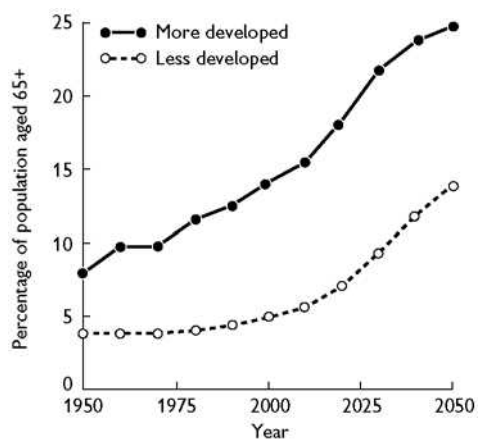


Fig. 8.3.1 Percentage of the population aged 65+ for more developed and less developed countries. (Reproduced from Kinsella, K. and Velkroff, V.A. The demographics of aging. *Aging Clinical and Experimental Research*, **4**, 59–69, copyright 2002 with permission from Editrice Kurtis S.r.l.)

Primary care

Unsurprisingly, the prevalence of depressive symptoms is considerably higher in elderly persons consulting their doctor than in the general community. One study in London found a point prevalence of about 30 per cent. Where it has been possible to compare the rates for those cases recognized by their doctor with cases independently ascertained by a research measure, such as a screening instrument or standardized interview, a typical finding is that the general practitioner recognizes about two thirds of the mild cases, rising to some 90 per cent of the moderate to severe ones. Some cases are considered to be depressed when they are not. Another finding is that elderly persons with depressive symptoms may not mention them to their doctor, attributing them to their age and circumstances. These findings have led to programmes offering additional training for GPs and to recommending the use of brief screening tests in primary care. Because diagnosis would lead to appropriate treatment being given earlier in an episode of depressive disorder, its duration would be shortened. Since prevalence is the product of incidence and duration, the prevalence of depression would therefore be expected to fall. This is an example of the application of epidemiology to prevention, which is its ultimate service.

Depression in hostels and nursing homes

Prevalence rates are also higher in hostels and nursing homes. In the United States, levels as high as 30 to 50 per cent have been reported. It might be thought that the context of living in a nursing home would account for having depressive symptoms. But one study found that the excess over the general population rates was largely accounted for by medical disorders, environmental factors contributing little to the variance. This needs to be studied further because it seems counterintuitive that the social and physical environment, both of which can be modified, could be of little relevance.

What is important is that only about one quarter of cases are recognized. To compound the situation, those cases that are recognized tend to be treated with too low doses of antidepressants. Depressive symptoms are well known to occur comorbidly with cognitive decline and the dementias. These findings from clinical epidemiology have pointed to the need for better case recognition through education of medical and nursing staff, and to the use of routine screening of residents in such settings.

Suicide

For many decades across the world, the traditional pattern has been for the highest rates of suicide to be in elderly men. This has now changed. In over a third of countries, both developed and less developed, it is younger people who have come to carry the highest rates. The World Health Organization provides a valuable resource for such data, showing rates by age and gender for nearly all countries from 1950 onwards.[6] The pattern varies considerably between countries. For example, in the United Kingdom in 2002, men aged 75 and over had a rate of 10 per 100 000, whereas the highest rate was 18 per 100 000, in men aged 35–44 years. In the United States and in the Russian Federation, the rates for men aged 75 and over were 41 and 89 per 100 000, respectively.

The main risk factors for suicide in the elderly are a past history of an attempt, depressive disorder, physical illness or disability, chronic pain, recent losses, social isolation, and access to lethal means. While universal interventions are more powerful than selected factors in prevention,[7] these attributes can be used in selective intervention to identify groups at increased risk. Furthermore, being multiplicative, these markers are of great value in individual cases by alerting the clinician to a person needing particular attention. Here is another example of the use of epidemiology for prevention. A systematic review of suicide prevention strategies for all age groups concluded that two interventions did reduce rates: physician education in recognizing and treating depression; and restricting access to lethal means.[8] Both of these interventions have close relevance to the elderly.

Personality disorders

The subject matter here refers to older people who have enduring attitudes and behaviour that bring difficulties for themselves or for others.[9] There is only sparse information on the prevalence of personality disorders in the general population, let alone specifically in the elderly. One exception, based on a national survey of mental health, found a lifetime prevalence of 6.5 per cent across all age groups with a trend towards lower rates with increasing age.

In clinical practice, it has long been suggested that traits such as impulsivity and externalizing behaviours tend to become less frequent in later life, whereas anxiety-prone, dependent, schizoid, paranoid, or obsessional persons are likely to change little as they age, or to become more so. Bergmann's pioneering enquiries among the elderly of Newcastle upon Tyne found that it was the anxiety-prone and insecure types that had late-onset neurotic disorders. A more recent study of late-life depression found an overall prevalence of comorbid personality disorder of 10–30 per cent. The group formerly known as neurotic and more recently as Cluster C in the DSM classification, had the higher prevalence. The Cluster B group, those with borderline, narcissistic, histrionic, and antisocial traits, were rare. What is not yet established, however, is if this lower prevalence also exists in the general population of the elderly, not just among cases with depressive disorder who have reached treatment in specialist services.

The epidemiology of personality disorders in later life is therefore significant for two reasons. First, some types are associated with increased risk of anxiety, depression, or paranoid states (*vide infra*). Second, there remains much yet to understand about the natural history of the personality disorders across the lifespan.

Psychosis of late onset

For the functional psychoses of late life, epidemiological information comes from two sources: studies of persons who have reached psychiatric services; and surveys of elderly persons living in the general community.[10] Psychotic symptoms probably exist as a continuum of severity, with only the more developed cases meeting diagnostic criteria. These often, but not always, reach psychiatric services, not uncommonly through being brought to the attention of the police. States phenomenologically similar to those found in clinics do occur in the community in non-trivial numbers. For cases that reach the threshold for a diagnosis by virtue of the range and severity of symptoms and behaviour, it has been proposed that cases with onset after the age of 60 years be called 'very-late-onset schizophrenia-like psychosis'. The syndrome has a 1-year prevalence

of 0.1 to 0.5 per cent. For advancing knowledge about the aetiology of schizophrenia, any information on it might be useful in explaining why people with this syndrome have reached the seventh decade or later in life without becoming psychotic, and only then develop it. It is more common in women. This is unlikely to be due to different social visibility or access to services. It is associated with a better premorbid level of social and occupational functioning. Premorbid paranoid or schizoid traits have been implicated and both clinical and community-based studies have found an association with sensory impairment such as deafness or poor eyesight. Personal and environmental factors associated with ageing have been considered, such as physical ill health, bereavement, loss of friends, and loss of income, but these have not been shown to contribute significantly. Genetic factors appear to be less important than in earlier onset schizophrenia.

Alcohol and drug dependence

It is generally believed that the prevalence of alcohol abuse and dependence declines during adult life and that the elderly have low rates in most communities. This may well be the case, but some other factors have to be considered. Whatever the prevalence, the absolute numbers will rise in the future because of the unprecedented growth in the elderly population. Next, the assumption may be false. In community surveys, errors in the ascertainment of alcohol abuse may lead to an underestimate for older persons. Most screening instruments were developed for use on younger adults, so their validity in the elderly is largely undetermined. Measures of the quantity drunk may mislead because smaller amounts may have an intoxicating effect in persons whose body fat, lean tissue, cerebral reserve, and metabolic function have declined. So the usual cut-off for problem drinking may be set too high for the elderly. One review of screening instruments concluded that the CAGE and MAST-G scales were appropriate, whereas other widely used instruments were not.[11] Next, all the studies have been cross-sectional. The elderly may have lower rates because of a cohort effect, whereby people born in the first half of the twentieth century may have been more moderate drinkers for all their life, compared to the high levels of consumption that are now found in the young of both sexes.

The actual values for prevalence are dependent on the instrument used and the definition used to define problem drinking, alcohol abuse, or dependence.[12] One review of community studies gives a figure of 5.1 per cent using various definitions. Invariably, men have higher rates than women. There is also considerable variation between countries and across different cultures. In identifying cases, a distinction of clinical significance needs to be made between late-onset and long-standing alcohol abuse. In primary care, accident and emergency departments, hospital in-patients, and nursing homes, the prevalence is much higher, yet cases are consistently under-recognized. The use of screening instruments in all of these settings has been advocated to improve this.

Alcohol abuse carries important comorbidity. In addition to all the established medical complications, it is associated with falls, subclinical delirium, cognitive decline, and depression. One study demonstrated a five-fold increase in the risk of developing a psychiatric disorder, especially depression and dementia. Simultaneous use of benzodiazepines, itself common in older persons, is clearly an additional and important factor. Against all this, it should be recalled that moderate alcohol use has been found in population studies to be associated with better mental and cardiovascular health, as well as being subjectively enjoyable.

Alzheimer's disease and other dementias

In the last two decades the dementia field has registered a tremendous scientific progression in many research areas including aetiology, pathogenesis, clinical aspects, treatment, and prevention. These advances have opened new perspectives, especially concerning definitions and diagnostic criteria, which have a relevant impact on epidemiological research.

Dementia is still defined as a syndrome which includes memory deficits and disturbance of other higher cortical functions; these major symptoms are commonly accompanied, and occasionally preceded, by deterioration in emotional control, social behaviour or motivation. However, it has become apparent that memory impairment may not necessarily be the major or first symptom for dementia subtypes such as frontotemporal dementia (FTD) and vascular dementia (VaD). Furthermore, as the current definition requires impairment severe enough to interfere with daily functioning, in several cases a delay of the diagnosis occurs. For that reason, a new research line has emerged with the aim to detect early Alzheimer's disease (AD) and other dementias, and the terms mild cognitive impairment (MCI) and cognitive impairment no-dementia (CIND) have been proposed to identify those subjects that show a clear cognitive deficit but do not fulfil diagnostic criteria for dementia. Finally, it is well known that dementia syndrome can be induced by many different underlying diseases, and that a differential diagnosis may be difficult for several reasons. AD as well as other dementia subtypes shows heterogeneity with distinct clinical and pathological characteristics; many different dementing disorders overlap in clinical and pathologic features; and different dementing disorders may make a common contribution or interact in causing dementia symptoms. Thus, rather than viewing, for example, AD and VaD as dichotomous entities, it may be more relevant to consider the role of their additive or synergistic interactions in producing a dementia syndrome.[13,14]

Following these new perspectives, in this chapter we will summarize the major findings from the most recent epidemiological research according to three major topics: early detection of AD and other dementias, incidence and risk factors for AD and dementia, and prevalence and impact of the dementing disorders at the individual and societal levels.

Early detection

As diagnostic criteria for AD require gradual onset of cognitive deficits, it is expected that cognitive disturbances are present already before the diagnosis can be rendered. Cognitive deficits are observable up to 10 years before dementia diagnosis with a sharp decline more evident in the final 3 years,[15] and occurring in episodic memory as well as in other cognitive domains such as executive functioning, verbal ability, visuospatial skills, attention, and perceptual speed.[16] However, our capability to use such early disturbances as a predictive tool of incipient dementia is strongly limited by several concomitant facts: (1) cognitive decline is also present as a function of the normal ageing process; (2) several conditions other than AD may lead to cognitive disturbances in the elderly; and (3) dementia-free patients with cognitive impairment observed in specialized clinical settings are different from cognitively impaired persons detected in the general population.[17]

To overcome these difficulties, different definitions have been proposed, with MCI and CIND being the most commonly used. MCI definition was originally derived in a clinical setting to identify subjects with isolated memory loss (now referred to as the 'amnestic' type) who may be in a preclinical phase of AD. Since then, the view has widened to cover a broader range of cognitive disturbances, and other MCI subtypes have been proposed.[18,19] CIND is derived essentially from population-based studies, and operationalized in slightly different items. Unfortunately, not one of the proposed definitions has shown a sufficiently good predictivity at the community level. Even a highly selected algorithm including subjective memory complaints, and global, and specific (memory/language) cognitive deficits could identify only 18 per cent of the incipient AD cases.[20,21] Although elderly persons with cognitive impairment have a high risk of developing dementia with a rate of about 11 to 50 per cent over 1 to 5 years, not all persons with CIND or MCI develop dementia. A substantial proportion of these persons (24–42 per cent) even improve in their cognitive performances over time.[22] This diverse prognosis supports the notion that AD is not the only causal mechanism underlying cognitive impairment in the non-demented elderly population. Cognitive psychologists have detected lower cognitive performances in elderly persons with deficiency in vitamin B_{12} and folate, elevated homocysteine, thyroid stimulating hormone deficiency, and cardiovascular disease. Physicians found an association between CIND and a number of factors including frailty-related factors such as history of hip fracture and high consumption of multiple drugs, history of psychoses, and depressed mood occurring 3 years before CIND development.[23] Other studies have identified older age, low education, depression, APOE ε4 allele, medicated hypertension, mid-life elevated serum cholesterol, and high diastolic blood pressure, as well as diabetes and anticholinergic medication use as risk factors for MCI.[17]

The prevalence of cognitive impairment—no dementia varies depending on diagnostic criteria with a maximum of 30 per cent.[24] In the younger elderly (e.g. 65–75 years old) cognitive impairment is actually more frequent than dementia disorders.[25] Annual incidence rates vary from 15 per 1000 among persons aged 75–79 years to 98 per 1000 among nonagenarians, when the estimates are corrected for dropouts due to death (Fig. 8.3.2; Ref.[26]). Regardless of the aetiology underlying the cognitive deficits, the high prevalence and incidence highlight the importance of this syndrome in the ageing population. Given that the criteria for cognitive impairment in non-demented persons are still under construction, and that there is no efficacious treatment to stop the possible progression to AD, at the moment we must be cautious with diagnosing MCI, which may cause an unnecessary burden on patients and relatives due to the unclear prognosis. However, it is clinically relevant to identify all persons with cognitive deficits due to treatable conditions such as depression, low-level vitamin B_{12}, and use of drugs.

Incidence and risk factors

Dementia incidence is similar in all continents and from different regions of the world (Table 8.3.1). Slightly lower rates detected in US in comparison with those in Europe and Asia are likely to be due to differences in study designs and case ascertainment.[27] AD accounts for 60–70 per cent and VaD accounts for 15–20 per cent of all dementia cases. The incidence of both AD and dementia increases almost exponentially with age. However, there are inconsistent findings regarding whether the rates continue to increase even in more advanced ages. The apparent decline found in some studies may be an artifact of the poor response rates, survival effects, and nature of populations previously sampled in these very old age groups.[28]

Age is the strongest risk factor for dementia and AD, suggesting that ageing-related biological processes could be implicated in the etiopathogenesis of AD. Further, the strong association with increasing age can be, at least partially, explained by a lifetime cumulative risk to different risk factors. Using this approach, the risk of dementia in late life is considered as a result of complex

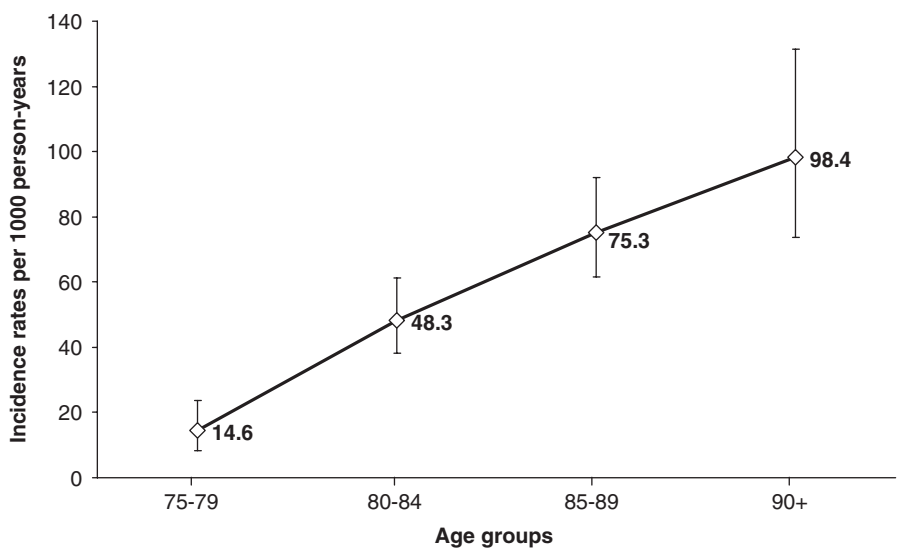


Fig. 8.3.2 Corrected age-specific incidence rates with 95 per cent confidence intervals of non-dementia cognitive impairment, including the two mutually exclusive definitions of amnestic MCI and CIND. (Reprinted from Alzheimer's and Dementia, 2, L., Fratiglioni, C. Qui, and K. Palmer, vascular cognitive impairment: time for prevention?, 202–4, copyright 2006, with permission from Elsevier.)

Table 8.3.1 Age-specific prevalence and incidence rates for dementia in the world and different regions: estimated from meta-analyses

Age groups	Incidence rate (per 1000 person-years)				Prevalence (per 100 population)			
	Worldwide (Gao *et al.* 1998)	Europe (Fratiglioni *et al.* 2000)	USA (Jorm and Jolley, 1998)	East Asia (Jorm and Jolley, 1998)	Worldwide (Jorm *et al.* 1987)	Worldwide (Fratiglioni *et al.* 1999)	Europe (Lobo *et al.* 2000)	China (Liu *et al.* 2003)*
60–64	1.1	—	—	—	0.7	0.9	—	0.3
65–69	3.3	2.4	2.4	3.5	1.4	1.6	0.8	0.7
70–74	8.4	5.5	5.0	7.1	2.8	3.5	3.0	1.3
75–79	18.2	16.0	10.5	14.7	5.6	6.9	5.8	2.8
80–84	33.6	30.5	17.7	32.6	11.1	13.0	12.0	5.6
85–89	53.3	48.6	27.5	72.1	23.6	25.2	17.4	11.8
90–95	72.9	70.2	—	—	—	35.8	28.5	23.7
95+	86.8	—	—	—	—	48.1	—	—

* The dementia cases in this meta-analysis include only Alzheimer's disease and vascular dementia.

(Reproduced from Backman, L., Small, B.J. and Fratiglioni, L. Cognitive deficits in preclinical Alzheimer's disease: current knowledge and future directions. In *New frontiers in cognitive aging* (eds. R.D. Dixon, L. Backman and L.G. Nilsson), pp. 161–77, copyright 2004, with permission from Oxford University Press.)

interactions of genetic susceptibility, biological factors, and environmental exposures experienced over the lifespan. A summary of all these factors is reported in Table 8.3.2, according to different biological mechanisms and grade of scientific evidence. As a part of an initiative of the Swedish Council on Technology Assessment in Health Care, specific criteria to summarize the scientific evidence concerning risk and protective factors for dementia have been proposed. Similar to criteria adopted for other diseases, these criteria first integrate the internal validity with basic causal criteria to weight the study quality, and then they take into account also number and proportion of the included studies reporting a specific association.[29] Moderate or strong evidence supports several genetic, vascular, and psychosocial factors as significantly related to both AD and dementia risk. Whereas implementing preventive strategies targeting the genetic susceptibility is limited, the other two hypotheses can easily lead to prevention programmes.

(a) Genetic hypothesis

First-degree relatives of AD patients have a higher lifetime risk of developing AD than the general population or relatives of non-demented subjects. Both genetic and environmental factors contribute to the phenomenon of familial aggregation. Twin studies have shown that heritability of AD is about 58 per cent, whereas other variance may be attributable to non-genetic factors.[30] The APOE ε4 allele is the only established susceptibility gene factor for both early- and late-onset AD. The risk effect of APOE ε4 allele decreases with increasing age, and after age 75 years, 15–20 per cent of AD cases are attributable to APOE genotype.[31] Familial aggregation of dementia and AD can be only partially explained by APOE polymorphism, implying that other genetic factors may be active and need to be detected.[32] Several other genes have been examined as possible candidates, but the reports are sporadic or the results are inconsistent.[33]

(b) Vascular hypothesis

Whereas the reported association between vascular risk factors and dementia risk is expected, due to vascular dementia, several explanations have been proposed for the association between vascular risk factors and Alzheimer-type dementia: (1) coexistence of vascular factors and AD pathology in the elderly; (2) precipitating effect of cerebrovascular disease or interactive effect between Alzheimer-type and vascular lesions in the brain; and (3) misclassification of mixed dementia as AD. Even if the mechanisms are still not fully understood, prevention may be possible as most vascular risk factors and diseases are modifiable or amenable to prevention and treatment. Controlling high blood pressure in middle age, avoiding mid-life obesity, and appropriately treating diabetes are the major intervention actions. Some studies also show that people who maintain tight control over their blood glucose levels tend to score better on tests of cognitive function than those with poorly controlled diabetes. Indeed, borderline diabetes or impaired glucose tolerance is also linked to an increased risk of dementia and AD in very old people.[34] Finally, to postpone clinical expression of the dementia syndrome in old people, preventing recurrent cerebrovascular disease as well as maintaining sufficient cerebral perfusion by adequately managing heart failure and avoiding very low blood pressure seems to be critical.

(c) Psychosocial hypothesis

Evidence from both epidemiological and biological studies indicates that factors acting at different periods across life course and having an intellectually stimulating nature may contribute in increasing the neural reserve and therefore promote functionally more efficient cognitive networks to cope with brain pathology and delay the onset of clinical manifestations of dementia. These factors include education, adult-life occupational work complexity as well as late-life social network and intellectually stimulating activities.[29,35] Although physical exercise may reduce the risk of brain damage due to atherosclerosis, the relevance of physical activity itself remains in debate, as most physical activities include also social and mental components. Complex leisure activities with physical, mental, and social components seem to have the most beneficial effect.[36] In addition to the reserve hypothesis, other mechanisms such as premorbid cognitive ability, vascular damage,

Table 8.3.2 Scientific evidence supporting risk and protective (in italic) factors of dementia and AD by different aetiological hypotheses.

	Risk and *protective* factors			
	Vascular	**Psychosocial**	**Genetic**	**Others**
Insufficient or limited scientific evidence	High cholesterol Cigarette smoking Obesity Late life high BP Late life low BP Heart failure Silent stroke *Moderate alcohol intake* *Dietary factors (e.g. fish and vegetables)*	Depression Low SES *Midlife physical activity* *Late life social network*	Several susceptibility genes	Head trauma Inflammatory markers *NSAIDs* *HRT* Folate and vitamins B12, A, E, and C deficiency Occupational exposure to toxics *Antioxidants*
Moderate or strong scientific evidence	Midlife high BP Diabetes mellitus Clinical stroke Atherosclerosis *Antihypertensive drugs*	Low education *Late life mentally stimulating activities* *Physical activity*	APOE ε4 allele* Familial aggregation	

Abbreviations: AD = Alzheimer's disease; APOE = apolipoprotein E gene; BP = blood pressure; HRT = hormone replacement therapy; NSAIDs = non-steroidal anti-inflammatory drugs; SES = socioeconomic status.

(Reproduced from Backman, L., Small, B.J. and Fratiglioni, L. Cognitive deficits in preclinical Alzheimer's disease: current knowledge and future directions. In *New frontiers in cognitive aging* (eds. R.D. Dixion, L. Backman and L.G. Nilsson), pp. 161–77, copyright 2004, with permission from Oxford University Press, New York)

neuroprotection, or detection bias may be possible explanations. The most likely effect of a mentally, physically, and socially active life is to postpone the onset of clinical dementia: even delaying dementia onset by 5 years would halve dementia prevalence and substantially decrease the number of dementia cases in the community.

Prevalence and impact

Despite different inclusion criteria, several meta-analyses of prevalence studies have resulted in strikingly similar results (Ref.[27]; Table 8.3.1). Currently, more than 24 million people in the world have dementia and this number will double in 20 years.[37]

The prognosis of dementia is dramatic. In 3 years, more than 50 per cent of the dementia cases progress to the severe stage. In the Kungsholmen Project, the proportion of severe dementia among prevalent cases increased from 19 per cent at baseline to 48 per cent after 3 years, and to 78 per cent after 7 years. This progression is due to both cognitive and functional decline. The mean annual rate of cognitive decline as measured with the MMSE varies from –4.0 to –2.0 points. Many predictors of a more rapid cognitive decline have been reported such as initial higher cognitive function, functional disability, and brain lesions. The APOE ε4 allele seems not to act as relevant prognostic factor.[38]

Dementia is strongly associated with disability being the major determinant of developing dependence and functional decline over 3 years. Approximately half of the persons who developed functional dependence in a 3-year period can be attributable to dementia.[39] In industrialized countries, mental disease and cognitive impairment are the most prevalent disorders among older adults living in nursing homes or other institutions. However, institutionalization of dementia patients varies depending on age structure, urban or rural residence, and other cultural aspects. In the 75+-year-old population, 70 per cent of incident dementia cases die during the 5 years following the diagnosis, accounting for a mortality rate specific for dementia of 2.4 per 100 person-years. Dementia triplicates the risk of death.[40]

Conclusion

Mental disorders are common chronic conditions among the elderly people, and the absolute number of subjects with psychiatric disorders will increase dramatically worldwide in the near future due to the ageing of the populations. In addition, the mental disorders have a high impact both at the individual and societal level. Prevention may represent one answer to these challenging conditions. The scientific advances of the last few years have provided sufficiently strong evidence supporting two possible preventative strategies: an active and stimulating lifestyle in late life as well as optimal control of other chronic disease both at middle and late age may decrease the risk of relevant psychiatric disorders such as AD and other dementias.

Further information

Hybels, C.F. and Blazer, D.G. (2003). Epidemiology of late-life mental disorders. *Clinics in Geriatric Medicine*, **19**, 663–96.

Blazer, D.G. and Hybels, C.F. (2005). Origins of depression in later life. *Psychological Medicine*, **35**, 1241–52.

Jorm, A.F. (2001). History of depression as a risk factor for dementia: an updated review. *The Australian and New Zealand Journal of Psychiatry*, **35**, 776–81.

Ferri, C.P., Prince, M., Brayne, C., *et al*. (2005). Global prevalence of dementia: a Delphi consensus study. *Lancet*, **366**, 2112–17.

Qiu, C., De Ronchi, D., and Fratiglioni, L. (2007). The epidemiology of the dementias: an update. *Current Opinion in Psychiatry*, **20**, 380–5.

Fratiglioni, L. and Wang, H.X. (2007). Brain reserve hypothesis in dementia. *Journal of Alzheimer's Disease*, **12**(1): 11–22.

Fratiglioni, L., von Strauss, E. and Qiu, C., (2007). Epidemiology of the dementias of old age. In *Oxford textbook of old age psychiatry* (eds. R. Jacoby, T. Dening, A. Thomas, and C. Oppenheimer), pp. 391–406. Oxford University Press, Oxford.

References

1. Kinsella, K. and Velkoff, V.A. (2002). The demographics of aging. *Aging Clinical and Experimental Research*, **4**, 59–69.

2. The US Centers for Disease Control and Prevention. (2003). Public health and aging: trends in aging—United States and worldwide. *The Journal of the American Medical Association*, **289**, 1371–3.
3. Henderson, A.S. (1994). Does ageing protect against depression? *Social Psychiatry and Psychiatric Epidemiology*, **29**, 107–9.
4. Klerman, G.L. and Weissman, M.M. (1989). Increasing rates of depression. *The Journal of the American Medical Association*, **261**, 2229–35.
5. Frombone, E. (1995). Depressive disorders: time trends and possible explanatory mechanisms. In *Psychosocial disorders in young people* (eds. M. Rutter and D.J. Smith). John Wiley & Sons, Chichester.
6. World Health Organization. (2007). Suicide prevention and special programmes. http://www.who.int/mental_health/prevention/suicide/country_reports/en/index.html.
7. Rose, G. (1993). Mental disorder and the strategies of prevention. *Psychological Medicine*, **23**, 553–5.
8. Mann, J.J., Apter, A., Bertolote, J., *et al.* (2004). Suicide prevention strategies: a systematic review. *The Journal of the American Medical Association*, **294**, 2064–74.
9. Burns, A., Bergmann, K., and Lindesay, J. (1998). Key papers in geriatric psychiatry. *International Journal of Geriatric Psychiatry*, **13**, 199–202.
10. Howard, R., Rabins, P.V., Seeman, M.V., *et al.* (2000). Late-onset schizophrenia and very-late-onset schizophrenia-like psychosis: an international consensus. The International Late-Onset Schizophrenia Group. *The American Journal of Psychiatry*, **157**, 172–8.
11. Beullens, J. and Aertgeerts, B. (2004). Screening for alcohol abuse and dependence in older people using DSM criteria: a review. *Aging and Mental Health*, **8**, 76–82.
12. Johnson, I. (2000). Alcohol problems in old age: a review of recent epidemiological research. *International Journal of Geriatric Psychiatry*, **15**, 575–81.
13. Agüero-Torres, H., Kivipelto, M., and von Strauss, E. (2006). Rethinking the dementia diagnoses in a population-based study: what is Alzheimer's disease and what is vascular dementia? A study from the Kungsholmen project. *Dementia and Geriatric Cognitive Disorders*, **22**, 244–9.
14. Fratiglioni, L., Qiu, C., and Palmer, K. (2006). Vascular cognitive impairment: time for prevention? *Alzheimer's & Dementia*, **2**, 202–4.
15. Bäckman, L., Small, B.J., and Fratiglioni, L. (2004). Cognitive deficits in preclinical Alzheimer's disease: current knowledge and future directions. In *New frontiers in cognitive aging* (eds. R.D. Dixon, L. Bäckman, and L.-G. Nilsson), pp. 161–77. Oxford University Press, Oxford.
16. Bäckman, L., Jones, S., Berger, A.K., *et al.* (2005). Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. *Neuropsychology*, **19**, 520–31.
17. Palmer, K. and Fratiglioni, L. (2006). Is mild cognitive impairment a distinct clinical entity? *Aging Health*, **2**, 763–9.
18. Gauthier, S., Reisberg, B., Zaudig, M., *et al.* (2006). Mild cognitive impairment. *Lancet*, **367**, 1262–70.
19. Portet, F., Ousset, P.J., Visser, P.J., *et al.* (2006). For the MCI working group of the European Consortium on Alzheimer's Disease (EADC). Mild cognitive impairment (MCI) in medical practice: a critical review of the concept and new diagnostic procedure. *Journal of Neurology, Neurosurgery, and Psychiatry*, **77**, 714–18.
20. Palmer, K., Bäckman, L., Winblad, B., *et al.* (2003). Detection of Alzheimer's disease and dementia in the preclinical phase: population-based cohort study. *British Medical Journal*, **326**, 245.
21. Sacuiu, S., Sjogren, M., Johansson, B., *et al.* (2005). Prodromal cognitive signs of dementia in 85-year-olds using four sources of information. *Neurology*, **65**, 1894–900.
22. Palmer, K., Wang, H.X., Bäckman, L., *et al.* (2002). Differential evolution of cognitive impairment in nondemented older persons: results from the Kungsholmen project. *The American Journal of Psychiatry*, **159**, 436–42.
23. Monastero, R., Palmer, K., Qiu, C., *et al.* (2007). Heterogeneity in risk factors for CIND. A population-based longitudinal study from the Kungsholmen project. *The American Journal of Geriatric Psychiatry*, **15**, 60–9.
24. Panza, F., D'Introno, A., Colacicco, A.M., *et al.* (2005). Current epidemiology of mild cognitive impairment and other predementia syndromes. *The American Journal of Geriatric Psychiatry*, **13**, 633–44.
25. De Ronchi, D., Berardi, D., Menchetti, M., *et al.* (2005). Occurrence of cognitive impairment and dementia after the age of 60: a population-based study from Northern Italy. *Dementia and Geriatric Cognitive Disorders*, **19**, 97–105.
26. Caracciolo, B., Palmer, K., Monastero, R., *et al.* (2008). Occurrence of cognitive impairment and dementia in the community: a 9-year long prospective study. *Neurology*, **70**, 1778–85.
27. Fratiglioni, L., von Strauss, E., and Qiu, C. (2007). Epidemiology of the dementias of old age. In *The Oxford textbook of old age psychiatry* (eds. R. Jacoby, T. Dening, A. Thomas, and C. Oppenheimer), pp. 391–406. Oxford University Press, Oxford.
28. Matthews, F. and Brayne, C. (2005). Medical Research Council Cognitive Function and Ageing Study Investigators. The incidence of dementia in England and Wales: findings from the five identical sites of the MRC CFA study. *PLoS Medicine*, **2**, e193.
29. Fratiglioni, L. and Wang, H.-X. (2007). Brain reserve hypothesis in dementia. *Journal of Alzheimer's Disease*, **12**(1), 11–22.
30. Gatz, M., Reynolds, C.A., Fratiglioni, L., *et al.* (2006). Role of genes and environments for explaining Alzheimer disease. *Archives of General Psychiatry*, **63**, 168–74.
31. Qiu, C.X., Kivipelto, M., Agüero-Torres, H., *et al.* (2004). Risk and protective effects of APOE gene towards Alzheimer's disease in the Kungsholmen project: variation by age and sex. *Journal of Neurology, Neurosurgery, and Psychiatry*, **75**, 828–33.
32. Huang, W.Y., Qiu, C.X., von Strauss, E., *et al.* (2004). APOE genotype, family history of dementia, and Alzheimer disease risk: a 6-year follow-up study. *Archives of Neurology*, **61**, 1930–4.
33. D'Introno, A., Solfrizzi, V., Colacicco, A.M., *et al.* (2006). Current knowledge of chromosome 12 susceptibility genes for late-onset Alzheimer's disease. *Neurobiology of Aging*, **27**, 1537–53.
34. Xu, W., Qiu, C., Winblad, B., *et al.* (2007). The effect of borderline diabetes on the risk of dementia and Alzheimer's disease. *Diabetes*, **56**, 211–16.
35. Fratiglioni, L., Paillard-Borg, S., *et al.* (2004). An active and socially integrated life in late life might protect against dementia. *Lancet Neurology*, **4**, 3343–53.
36. Karp, A., Paillard-Borg, S., Wang, H.X., *et al.* (2006). Mental, physical and social components in leisure activities equally contribute to decrease dementia risk. *Dementia and Geriatric Cognitive Disorders*, **21**, 65–73.
37. Ferri, C.P., Prince, M., Brayne, C., *et al.* (2005). Global prevalence of dementia: a Delphi consensus study. *Lancet*, **366**, 2112–17.
38. Agüero-Torres, H., Qiu, C., Winblad, B., *et al.* (2002). Dementing disorders in the elderly: evolution of disease severity over 7 years. *Alzheimer Disease and Associated Disorders*, **16**, 221–7.
39. Agüero-Torres, H., Fratiglioni, L., Guo, Z., *et al.* (1998). Dementia is the major cause of functional dependence in the elderly: 3-year follow-up data from a population-based study. *American Journal of Public Health*, **88**, 1452–6.
40. Agüero-Torres, H., Fratiglioni, L., Guo, Z., *et al.* (1999). Mortality from dementia in advanced age: a 5-year follow-up study of incident dementia cases. *Journal of Clinical Epidemiology*, **52**, 737–43.

8.4

Assessment of mental disorder in older patients

Robin Jacoby

The assessment of older people is not fundamentally different from that of younger patients. The principles of taking history and mental-state examination are the same at any age. But if the goals are common, the routes taken to reach them are not necessarily so. For example, an assessment adequate enough to begin treatment of a 30-year-old woman presenting to an outpatient clinic with a depressive illness might take about an hour and involve speaking only to the patient and perhaps briefly to her partner, whereas the equivalent assessment of an 81-year-old woman in whom uncertainty exists as to whether the diagnosis is that of a depressive or a dementing illness may require more than one interview and necessitate enquiry from several informants. This section will not repeat what can be found in Chapter 1.8.1, but cover only those points which are specific to or need to be emphasized for older patients.

The referral process

Who refers?

Whilst the referral process might be the same as for younger patients, it is more often different. In many cases the patient has no idea why, or indeed does not even know or has forgotten that she has been referred. (The feminine gender is used in this chapter because older women are more likely to develop a mental illness and to survive longer than men. However, what is written applies also to men.)

The process has most often been initiated by family members who might not have discussed it with the patient. Many old people live alone with no relatives nearby or even in the same country or state, so that referrals are frequently initiated by friends, neighbours, or other acquaintances, such as local shopkeepers, social services care workers, and people who run luncheon clubs.

Reasons for referral

In the case of a woman of 30 with a depressive illness, she is referred to a psychiatrist for treatment to effect a remission. However, an older woman of 80 with a similar condition may be referred for a variety of reasons including the following: the primary care doctor might be uncertain of the diagnosis, that is whether it could be dementia; the grown-up children might have removal from home to residential or nursing care as the first item on their agenda; the patient's condition may not be the primary issue—there may be greater concern for her husband who is failing to cope, perhaps to the extent of physically abusing her.

The informants

A large number of older people seen by psychiatrists are unable to give complete or reliable information about themselves. Frequently, but not invariably, there is a spouse or adult offspring living with the patient. In other cases, however, it is necessary to track down someone less obvious. Neighbours are often helpful at relating recent history, but may know little of past personal or family history. Effort spent in telephoning relatives, even those on the other side of the world, can be invaluable in giving an account of such items as family history or premorbid personality. If an informant is not readily available, for example, because it is night-time in Australia, the psychiatrist should not shelve the task of phoning, but only defer it to the next available opportunity.

Where conflicting information is given by a variety of informants it might be necessary to weigh up the particular 'hidden agenda' of each one. For example, the husband of a demented woman may minimize his wife's behaviour disturbance for fear that she would be 'put away'; whereas the daughter may overstate it in order to support a case for her mother's transfer to a nursing home because her father repeatedly phones her for assistance at all hours of the day and night. Each one of the two informants has cogent reasons for weighting the information, but the psychiatrist and his or her team cannot help to resolve the situation until they understand those reasons.

Professional informants

Psychogeriatrics is as dependent on multi-disciplinary working as any other branch of psychiatry. Many patients seen for the first time will already be well known to their primary care doctor who will be able to provide invaluable information. The same frequently applies to community psychiatric nurses who now take referrals directly from general practitioners and may themselves be making referrals to the old-age psychiatry service. The psychogeriatrician can save a great deal of time and effort by consulting community psychiatric nurses and general practitioners before seeing the patient or relatives.

Where to assess the patient

The patient needs to be placed at her maximum advantage to provide clinical information in whatever setting the assessment takes place. This has to be stated explicitly because the doctor is often required to take active steps to ensure it. Account has to be taken of special sensory impairment. Poor vision may need lights to be switched on so that the patient can see who is asking her questions. Distracting noises will make it even more difficult for someone with hearing impairment to grasp what is said. Surprisingly often, this may require a request that the television be switched off. Most importantly, examiners need to sit facing the patient with the lips visible, to speak slowly, and to enunciate words carefully. The patient should then be asked if she can hear properly. Simply shouting at her is not a substitute for these simple steps.

Social customs vary within and between societies. For instance, in the United Kingdom and the United States the use of first names is much more acceptable with younger adults than it was 40 years ago. With the current generation of older patients it is not. For them to be called by their first names unbidden is disrespectful and infantilizing. Even if nurses and other non-medical staff do so, doctors should not use first names, unless specifically invited. Instead, the surname plus appropriate title (Mr, Mrs, etc.) is correct.

At home

The preferred place to assess older patients is in their own homes, although circumstances sometimes dictate that it will be elsewhere. At home patients feel less intimidated and can be seen within an environment which tells the psychiatrist a great deal that he cannot know in the clinic. If a house is filthy and cold and the patient in a similar state, and if there is reliable information that this is only a recent phenomenon, then it is a powerful descriptor of the patient's inability to cope. However, the converse is not always true; a clean and tidy home may only reflect someone else's willingness to support and care for the patient who could not otherwise do it herself (e.g. a daughter or neighbour). Another advantage of a home assessment is that cognitive disabilities, such as dyspraxia and agnosia, can be tested in an ecologically valid way (making tea, recognition of family members from photographs) that is more acceptable to a patient than being formally tested with the Mini-Mental State Examination.[1]

Assessments at home require more preparation for the doctor than is necessary at outpatient clinics where equipment for physical examination and blood tests are available, for example. It is an obvious courtesy to the patient to let her know of the visit beforehand, but it is also wise to arrange for a suitable informant to be present. Furthermore, some older patients are incapable of letting visitors into their houses and the informant might well first have to facilitate the doctor's access. Elderly patients are much more likely to be suffering from comorbid physical illness which may be the fundamental cause of the mental disorder, for example, pneumonia or a urinary tract infection manifesting delirium. The old-age psychiatrist does not therefore need to adhere rigidly to lines of specialty demarcation but rather be aware of the possibility of and prepared to search for physical illness. The basic equipment for a medical examination, such as a stethoscope, sphygmomanometer, and patellar hammer are items to be taken on home visits. Urine testing strips and a thermometer, especially a low-reading thermometer, are also sometimes useful.

In a psychiatric hospital

Patients who are assessed after admission to psychiatric beds lack the advantages of being in their own environment, although the opportunity for physical examination is much easier. Another advantage for hospital inpatients is that the assessment can be carried out over a longer period of time, since older people tire more easily and cooperation varies from day-to-day. For example, some demented patients will object to undergoing full cognitive assessment in one go, especially because they are often aware that they are failing. If a few questions are asked in the course of several short sessions, a more accurate and complete picture of the patient's abilities eventually emerges. If the Mini-Mental State Examination is administered in this way, a higher total can be achieved than if an attempt to administer it all at once meets with sullen refusal after the first few questions, with all subsequent ones having to be scored zero.

Information from other informants is as crucial for hospital inpatients as it is for those seen at home. It is usually the responsibility of the house officer or resident to collect the history, and they may be required to telephone several informants in distant and local parts to obtain a full picture which the patient is incapable of providing.

Liaison visits in general hospitals

Liaison visits to patients in general hospitals make up a considerable part of the old-age psychiatrist's work because comorbid mental and physical illnesses are very common. In spite of the fact that the host nurse's instinct is to lead the visiting psychiatrist straightaway to the patient's bedside, the latter should insist on first reading the case notes (charts) and speaking to the nursing staff who know her best. From the case notes and the prescription cards (medication orders) invaluable information on current and past drug therapy as well as details of the patient's medical history are obtained. Clues as to the patient's mental state are often best gleaned from the records written by the nurses. Nevertheless, non-psychiatrist doctors, surgeons, and nurses are not accustomed to assessment of the mental state and statements such as 'confused' should not be taken at face value, since they stand for anything from slight difficulty in answering complex questions due to anxiety at being in a strange environment to major mental disturbance. As in most other settings, time spent telephoning informants from the general hospital ward is well invested and may permit the visiting old-age psychiatrist to express an opinion on the patient's condition more firmly than would otherwise have been possible.

A useful final step before going to talk to the patient is, if possible, to observe her from a suitable distance. In this way signs of delirium, disruptive behaviour, social interaction, and other phenomena such as dyskinesias may be seen.

When seeing the patient herself, wherever possible she should be taken to a separate room and not examined in an open ward where there are other patients. If it is impossible for the patient to leave her bed, then it is usually feasible to move the bed to a more private place.

Nursing and residential homes

Much of that which is required for liaison visits to general hospitals applies to assessment in residential or nursing homes, most notably trying to see the patient in a private room away from other residents. Since abuse of elderly people is sometimes an issue in these

settings, it is preferable to have at least some time completely alone with the patient first, and if indicated, to check for bruises or other injuries, and secondly to allow the patient to tell the doctor things which she might be frightened to do in front of the staff of the home. Another problem in some nursing and residential homes is that the psychiatrist finds that an untrained or unqualified member of staff accompanies him, the quality of whose information may not be at the level of trained nurses. Careful questioning of several members of staff, attention to written records, and telephone calls to appropriate informants should all improve the quality of the assessment.

The history

Family and personal history

As has already been made clear, for many older patients a complete history may have to be obtained from a variety of informants. With the patients themselves a more flexible approach than is taken with younger ones is often needed. Whether intellectual failure is obvious and global or there is only relatively mild cognitive impairment (MCI), for some to give a history that is fully chronologically correct can be too great an effort. The examiner must accept these limitations and try to keep the atmosphere as relaxed as possible. Much more than the young, elderly psychiatric patients perceive the psychiatric interview as an ordeal or a form of trial in which it is easy for them to acquire a sense of failure. This in turn induces anxiety and a vicious spiral of ever worsening performance. One way in which the patient can be put at ease is to reassure her that you will come to her main problem in due course but that it would be good to hear something of her background first. For most older patients the family and personal histories are easier to recall than the confusing events which have led up to the referral. This is not simply good for the patient but for the examiner as well. Amongst the most profitable of pleasures in old-age psychiatry are the life stories of people who have lived during some momentous periods of world history. Furthermore, these stories put patients into a context which makes it much easier to understand why and how they have reacted to the mental illness with which they have presented.

As regards the family history specifically, the examiner needs to be alert to mistakes which could indicate cognitive impairment. A patient may confuse family relationships or misidentify family members quite early in the course of a dementing illness. In other words, inaccurate information from the patient can be as clinically informative as that which is correct. Older patients are not necessarily as sophisticated in medical vocabulary as their children and grandchildren. Therefore, to obtain facts about a possible family history of dementia (an important issue), it may be necessary to ask if any blood relative had 'memory problems' in late life or 'had to go into a home'.

In eliciting the personal history the examiner might need to be aware of the historical context at the time in question. Some older patients, however affluent they may be now, grew up in poverty or other adverse circumstances (e.g. a parent died of tuberculosis during their childhood) which could still be affecting their psychological lives. Similarly, education may have been disrupted in a way that is more unusual nowadays. Some patients, notably women, relate how they missed education because they had to look after their younger siblings after mother died or father was killed, because the remaining parent had to go out to work for the family to survive. A precise enquiry should be made as to educational attainment, and especially the level of literacy and numeracy which may be the only 'baseline' appraisal obtainable for a patient with current dyslexia or dyscalculia.

Whilst it is often very obvious to the examining doctor that a patient has cognitive impairment because of her errors and inconsistencies in giving her personal history, it is very unwise to foreclose on a diagnosis at this stage. Patients with severe depressive illnesses can be even more hesitant than those with, for example, Alzheimer's disease, or show such retardation and lack of concentration on what is being asked that they can portray a clinical picture that is not easily distinguishable from a state of advanced dementia.

Medical and psychiatric history

The nature of the information required is no different from that in younger patients. However, of particular importance in the older population is past and present medication. Drugs taken at the prescribed dose, at a wrong dose (due to dementia), or drugs no longer intended to be taken and prescribed sometimes quite a long time ago but of which a residual supply remains, are all potent causes of confusion and even frank delirium in old people. It is therefore good practice to ask to see where all medication is kept and to examine each pack or bottle to check that the amount left is approximately proportionate to that which one might expect given the date the drugs were dispensed. Since elderly people are frequently the victims of clinically injudicious polypharmacy, it is common to find a large quantity of current and prescription-expired drugs which the patient is taking on a random basis. To counter this problem proprietary boxes which dispense drugs in daily amounts, such as the Dosette or Nomad systems, are used and can more easily be checked for compliance or overdosing.

Premorbid personality

Personality is one of the prime determinants of outcome in mental illness at any age, but where older people are concerned too little effort is made, partly for lack of reliable informants, to give a valid assessment of personality. False assumptions are made that someone has always been awkward or cantankerous, whereas it is shown later that they have become so because of frontal-lobe impairment. Similarly, it is sometimes assumed that a woman has always exhibited attention-seeking or manipulative behaviour, when to the surprise of doctors and nurses alike such behaviour disappears following effective treatment for a depressive illness. Every effort should be made to find a reliable informant before reaching such conclusions.

The mental state examination

Appearance, behaviour, and the environment

A great deal can be learnt from the appearance of the patient and her home environment. Signs of neglect in both are commonly found. A brief tour of the home may reveal rotten food, little or no food, empty bottles of liquor, evidence of poor hygiene or incontinence, and inadequate heating. The patient may be dirty and unkempt, and clothes may have been put on in the wrong order (dressing dyspraxia). Particular attention should be paid to relatively mild impairment of attention and concentration, since it

might betray a delirium which can be treated to achieve a remarkable improvement in the patient's condition. Agitation is another sign to be carefully sought. Sometimes agitation is obvious with behaviour such as pacing and sighing which make it almost impossible to communicate with the patient, but at other times it can be allied to psychomotor retardation and perceptible only in tireless movements of the fingers.

Talk

Dysphasia in any of its guises is a frequent manifestation of dementia. Sometimes it is obvious, but not always. For example, it may be difficult at first to differentiate between an expressive dysphasia (Broca's dysphasia) and retardation. It is relatively mild or moderate receptive dysphasia (Wernicke's dysphasia), however, which traps the unwary clinician. In such cases the patient may appear to be obtuse, unintelligent, or hard of hearing until it is appreciated that she simply does not understand a considerable proportion of what is said to her.

Even if the patient is not dysphasic, she may be evasive and given to circumlocution. Again, this should not be taken automatically as evidence of a premorbid lack of intelligence or pompousness, since it is frequently used as a camouflage for cognitive impairment. A patient with dementia might say 'Oh of course I know that' or 'I never paid any attention to that sort of thing' when asked to give an item of current affairs.

In manic or hypomanic illness in old-age slow flight of ideas is sometimes missed by inexperienced clinicians or mistaken for evidence of cognitive impairment. Here the normal coherence of thought is disrupted because the patient is distracted from one idea to another, just as in characteristic flight of ideas, but they are delivered at a normal or even slower pace. The latter can occur if a mixed affective state is present.

Thought content

Much of that which applies to younger patients applies also to older ones and does not need to be mentioned here. However, subtle changes in thought content amounting to a restriction in breadth and a repetitiveness of themes may be noticed. Formal thought disorder which is found in young patients with schizophrenia is extremely rare in the old.

Mood

Depressive illness is commonly missed in older patients. This is partly because the clinical picture can mimic a dementing illness, so that a history from a reliable informant, as has already been stressed, is mandatory. Another reason is so-called *masked depression* in which the patient denies depressed mood but presents with other symptoms, such as those of apparent physical illness. In this sort of case a reliable account of sleep and appetite disturbance, weight loss, and anhedonia give clues as to the presence of an affective disorder. It should also be remembered that dementia and depressive illness are common disorders and not infrequently occur together, so that the diagnosis of one does not rule out the other.

Older men are still the group most at risk of suicide in most countries of the world where statistics are recorded (see Chapter 4.15.1). The psychiatrist does not shrink from specific enquiry about suicidal thinking in patients of any age, but it can sometimes be difficult to differentiate between a rational desire to die when the time comes and active suicidal ideation. In-depth probing is therefore mandatory and the examiner should not be put off by the patient's attempts to leave the topic, if he or she thinks that there is likely to be risk of self-harm.

Cognitive examination

In assessing elderly patients more emphasis is usually placed on the cognitive examination than in younger patients. In theory it can be as exhaustive and thorough as assessment by a neuropsychologist, but in the routine practice of old-age psychiatry it usually has to be feasible within the constraints of a consultation which lasts about an hour.

It is in this part of the assessment that the examiner is most likely to lose the patient's cooperation, principally because of the humiliation experienced by some at their own failures. Some patients become angry, indignant, or defensive. Others become anxious and their performance deteriorates. One way to pre-empt this is to preface testing by stressing that this is not a competitive examination and that most people have difficulty answering some of the questions. Correct answers are praised without excessive emphasis and incorrect ones are either treated in a neutral way or given a positive spin by saying, for example, 'Well it's…, but you weren't far off'.

Many old-age psychiatrists and other members of their multidisciplinary team prefer to use standardized questionnaires, such as the Mini-Mental State Examination,[1] the questionnaire in widest use. It has the advantage that results between and within patients can be compared and progress can be monitored. However, no off-the-shelf test is exhaustive and none produces an adequate cognitive assessment by itself. The clinician therefore needs to have some sort of schema for covering the main areas of cognitive function which would include: memory (in its various aspects) and general information, and naming; the understanding and production of language; praxis (ideomotor and constructional); sensory recognition (gnosis); abstract reasoning; verbal fluency; calculation; left/right orientation; and executive function (the ability to integrate mental processes for goal-directed activity). The list is not exhaustive and some areas may need to be covered in greater detail as the clinical situation demands. Table 8.4.1 gives a guide to cognitive examination based on an extended Mini-Mental State Examination. Clinicians vary in the order and way in which they test individual cognitive functions, and the list in Table 8.4.1 is not intended to be prescriptive.

Other aspects of the mental state examination

These do not differ in essence from that in younger adults and are not covered in this section.

Physical assessment

Since physical comorbidity is extremely common, the old-age psychiatrist needs to be able and willing to conduct a basic physical examination. In the patient's home this may not always be easy. For instance, it may not be kind to ask a frail person who takes an hour or more to dress and come downstairs in the morning to return to her bed and undress, but it is usually possible to make a reasonable examination of the arterial pulse, blood pressure, and jugular venous pulse, and to auscultate the heart and lungs when

Table 8.4.1 Schema for testing cognitive functions based on an extended mini-mental state examination

Function	Subfunction	Examples
Orientation	Time	Year, month, date, day, season
	Place	Own address or that of hospital, city, county/state, country
	Person	Own name (married women sometimes cannot give married name); recognize others by name or function (e.g. you are a doctor)
Memory	Immediate recall	Immediate repetition of three objects or a name and (local) address
	Delayed recall	Repetition as above but after a distractor task
	Long-term recall	Give historical or personal events (that can be verified)
	General information	Names of politicians or other VIPs
Concentration		Months of the year in reverse order, counting from 20 back to 1; spelling WORLD forwards then backwards
Praxis	Construction	Copy diagram of interlocking pentagons
	Ideomotor	Draw a clock and set the hands at a specified time (also a test of executive function)
	Dressing	Put on a jacket; undo, and refasten buttons
Sensory recognition (gnosis)	Visual including prosopagnosia	Recognize photographs taken from unusual angles and of familiar faces
	Auditory	Recognize the doorbell
	Tactile	Recognize objects placed in the palm, e.g. coins
	Reading	Any sample *but* use large print, e.g. newspaper headlines
	Olfactory	Recognize something from the kitchen, e.g. coffee
Language	Expressive	Repeat 'no ifs ands or buts'
	Understanding	Carry out a three-stage command
	Naming	Naming objects of increasing complexity
Verbal fluency		List as many items from a category as possible in 1 min, such as boys' and girls' forenames, or as many words beginning with a specified letter
Writing		Write an ordinary English sentence
Calculation		Not too complex—subtraction of serial sevens from 100 is too difficult for many. A simple sum involving money is better
Left–right orientation		Face–hand test (e.g. left hand to left ear, right hand to left ear); finger recognition on own and examiner's hand
Abstract reasoning		'In what way are an apple and a banana alike?'
		'In what way are a boat and a car similar?'
		Interpretation of simple proverbs

Testing one function usually depends on one or more others; for example, most tests depend on understanding of language. Examples are neither prescriptive nor exhaustive.

the patient is seated. Similarly, a partial neurological examination for signs of focal deficits is also possible. However, if something alerts the doctor to the need to examine an undressed and supine patient, the duty should not be shirked, lest a hitherto unsuspected abdominal mass, or a strangulated hernia are missed.

Patients admitted to psychiatric hospital or nursing home beds should all undergo a physical examination. Focal neurological signs may indicate the cause of dementia. Carcinoma of the breast which either dementia or fear has prevented the patient from declaring may be much more treatable than she has believed. All the physical disorders which may be revealed are too numerous to mention, but their detection and treatment nearly always contribute to an improvement in mental function.

Laboratory investigations

Owing to tight budgetary constraints it is sometimes argued that routine laboratory investigations, such as full blood count and chest radiography, are unnecessary for younger adults. Whether or not this is true, with older patients such tests are strongly advised because the treatment of comorbid physical illness improves mental disorder. Furthermore, a treatable or arrestable cause of dementia may be found. For patients with dementia the following are recommended: full blood count; serum electrolytes, and creatinine; liver and thyroid function tests; syphilis serology; vitamin B_{12} and red cell folate; chest radiography. Medical and nursing staff should also have a low threshold for sending urine for microbiological examination (see superimposed delirium below). The vexed question of neuroimaging, an expensive procedure, is much discussed and the debate is not easily summarized or resolved. Most space-occupying lesions can be detected, as can many vascular changes. However, there is nothing pathognomonic for Alzheimer's disease on CT, magnetic resonance imaging, or single-photon emission tomography. In Alzheimer's disease, scans may support the diagnosis but not establish it.

General considerations in the assessment of older psychiatric patients

Falls

The causes of falls in older people are many and the reader is referred to textbooks of geriatric medicine for a full discussion. However, the old-age psychiatrist needs to be aware that many

psychotropic drugs precipitate falls through postural hypotension, with tricyclic antidepressants and neuroleptics being particular offenders. Neuroleptics induce parkinsonism, putting the patient at risk of tripping against rugs or items of furniture. The clinical implications of falls, especially in older women, are serious because patients are at risk of a fractured neck of femur or (less commonly) a subdural haematoma, both of which carry a high mortality.

Nutrition

Poor nutrition is commonly found in patients presenting to old-age psychiatry services. Even in affluent societies many older people are amongst the most impoverished or feel that they cannot afford good food. Some are too frail to get out to the shops. Others lack motivation to shop and eat because of depressive illness. Patients suffering from dementia may be incapable of shopping and preparing food. Widowers might never have learned to cook. A vitamin B_{12} or folate level at the lower end of the reference range, or even below it, is as often a reflection of poor nutrition as an indication of pernicious anaemia.

Superimposed delirium

Here it is mentioned only that subacute delirium superimposed on another condition is at risk of not being recognized because it is taken to be a manifestation of the underlying illness, usually dementia. This is particularly the case when a subclinical urinary tract infection occurs in a demented patient. It is of clinical relevance because treatment of the urinary tract infection results in a great improvement in the patient's mental state. On routine assessment particular attention should therefore be paid in the history to evidence of sudden worsening of a stable or only slowly deteriorating condition, and to nocturnal disturbance especially with (usually visual) hallucinations. On examination of the patient herself the level of consciousness, awareness of the environment, attention, and concentration should be noted.

Delirium is described in Chapter 4.1.1, and special features in older people are considered in Chapter 8.5.1.

Reassessment after treatment

Because so much of psychiatry is practised in the community, it is impossible for all patients to be reassessed by a psychiatrist after treatment. In hospital or outpatient clinics it is feasible, but not for the majority of older patients who live at home. Furthermore, in areas where the population is geographically widespread, it is very difficult if not impossible for many older people, who may be frail and infirm, to travel long distances to attend clinics. Much of the follow-up assessment in old-age psychiatric services is therefore carried out by other members of the multi-disciplinary team, such as psychologists or occupational therapists, but mostly by community psychiatric nurses. It is essential for the psychiatrist to meet regularly with members of the team seeing patients in the community to discuss the progress of individual patients following treatment.

Further information

Goldberg, D. and Murray, R. (eds.) (2006). *The Maudsley handbook of practical psychiatry* (5th edn). Oxford University Press, Oxford.

Hodges, J.R. (2007). *Cognitive assessment for clinicians* (2nd edn).Oxford University Press, Oxford.

Jacoby, R., Oppenheimer, C., Dening, T. *et al*. (eds.) (2008). *The Oxford textbook of old age psychiatry*. Chap. 10–12. Oxford University Press, Oxford.

References

1. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189–98.

8.5

Special features of clinical syndromes in the elderly

Contents

8.5.1 Delirium in the elderly

James Lindesay

Note Dementia in people of all ages is considered in Part 4, Section 4.1, where the following topics are considered: Chapter 4.1.2 Dementia: Alzheimer's disease; Chapter 4.1.3 Frontotemporal dementias; Chapter 4.1.4 Prion disease; Chapter 4.1.5 Dementia with Lewy bodies; Chapter 4.1.6 Dementia in Parkinson's disease; Chapter 4.1.7 Dementia due to Huntington's disease; Chapter 4.1.8 Vascular dementia; Chapter 4.1.9 Dementia due to HIV disease; Chapter 4.1.10 The neuropsychiatry of head injury; Chapter 4.1.11 Alcohol-related dementia (alcohol-induced dementia; alcohol-related brain damage); Chapter 4.1.13 The management of dementia.

Introduction

Although delirium occurs at all ages, it is most frequently encountered in late life. This is because delirium is the result of an interaction between individual vulnerability factors (e.g. brain disease, sensory impairment) and external insults (e.g. physical illness, medication), the rates of which both increase with age. Our current concept of delirium derives principally from the florid clinical stereotype that has evolved from centuries of clinical observations on younger patients, and it may not be applicable to our historically unique ageing population. In younger adults, a major physical insult is usually necessary to precipitate delirium, which is often a dramatic disturbance. This is not the case in vulnerable elderly patients when relatively mild physical, psychological, or environmental upsets may be sufficient to bring about acute disturbances of mental functioning. These disturbances may be less obvious than in younger patients, particularly if they occur in the context of pre-existing cognitive impairment. Consequently, despite being common and problematic, delirium in elderly patients is frequently missed or misdiagnosed as dementia or depression by medical and nursing staff.[1] This is unfortunate, because delirium is an important non-specific sign of physical illness or intoxication, and if left untreated there may be costly consequences, both for the patient and for health services.

Clinical features

The clinical features of delirium are described in Chapter 4.1.1. Most delirium in elderly patients is of the quiet hypoactive variety, lacking the more florid disturbances in mood, perception, and behaviour that bring the disorder to clinical notice. Reversible cognitive impairment in elderly patients is associated with reduced conscious level, poor attention, poor contact with the patient, incoherent speech, reduced psychomotor activity, lack of awareness of surroundings,

poor orientation, and poor memory.[2] Hyperactive delirium does occur in elderly patients, but it is less pronounced, with the overactivity usually confined to purposeless behaviour such as pulling at the bedclothes. Violent behaviour is uncommon; elderly patients are more likely to injure themselves than others.

Classification

The ICD-10 and DSM-IV diagnostic criteria for delirium are described in Chapter 4.1.1. They are not entirely concordant; ICD-10 is more restrictive, resulting in the diagnosis of fewer cases.[3] However, the two systems agree on four essential features: disturbance of consciousness, disturbance of cognition, rapid onset/fluctuating course, and evidence of an external cause. Unfortunately, none of these features is specific for delirium as opposed to dementia, and the current diagnostic criteria are poor predictors of outcome, defined in terms of improvement in cognitive function. Reversibility of cognitive impairment may be the most discriminating feature of delirium,[2] but is problematic as a diagnostic criterion since outcome is unknown at the outset.

Another shortcoming of the current classifications of delirium is that they do not recognize the partial and transitory disturbances that are commonly observed in elderly patients. Subsyndromal delirium is common, and is part of a continuum between normality and the full syndrome. Subsyndromal cases are clinically significant, since they have the same risk factors and the same increased mortality as syndromal cases.[4]

Diagnosis and differential diagnosis

The diagnosis of delirium is a two-stage process: first, diagnose the delirium, and second, identify the underlying cause or causes. The diagnosis of delirium in elderly patients can be problematic, given the predominantly hypoactive clinical picture and the unreliability of 'positive' symptoms. However, it is important to consider the possibility if cognitive decline is rapid, and if any of the recognized signs and symptoms are present. A good informant history from relatives or ward staff is essential to establish the onset and course of the disorder. Routine screening procedures may be useful in identifying patients who develop delirium while in hospital. Brief instruments such as the Mini-Mental State Examination[5] are not diagnostic, but will alert the clinician to any sudden decline in cognitive function. More extended diagnostic instruments are also available, such as the Delirium Rating Scale,[6] the Confusion Assessment Method,[7] and the Delirium Symptom Interview.[8] Another approach to screening for delirium is to identify those at particular risk of developing the disorder. Predictive factors related to the patient include: visual impairment, severity of illness, cognitive impairment, and a blood urea nitrogen/creatinine ratio of 18 or more.[9] Hospital- and treatment-related factors include: use of restraints, malnutrition, use of more than three medications, bladder catheterization, and the number of iatrogenic events.[10] These factors are multiplicative in their effect.

The differential diagnosis of delirium includes most other psychiatric disorders in this age group. These disorders are themselves risk factors for delirium, so the possibility of co-morbidity must always be considered. When in doubt, investigate and manage as delirium until the situation is clear.

Dementia

Dementia is a major risk factor for delirium, and in practice co-morbidity commonly occurs. However, differential diagnosis is important, as episodes of delirium need to be identified in order for them to be managed effectively. Recent onset and rapid decline of cognitive functioning, from whatever baseline, indicate an episode of delirium until proved otherwise. Delirium in elderly patients can be prolonged, and failure to recover quickly following treatment of the cause does not necessarily indicate an underlying dementia. It is important to have a good history of pre-morbid functioning.

Depression

Delirium can be difficult to distinguish from severe depression in elderly patients, cognitive impairment associated with severe depression is usually relatively mild in comparison with the affective disturbance, whereas the reverse is true of delirium. The pattern of diurnal variation also varies in the two disorders, with depressed patients tending to be worse in the mornings, and delirious patients in the evenings. Elderly depressed patients are at increased risk of delirium, either through self-neglect or because of the antidepressant treatment they are receiving. Anticholinergic tricyclic drugs are particularly troublesome in this respect. Adverse life events, such as bereavement, may precipitate both depression and delirium in vulnerable individuals.

Mania

Mania is much less common than delirium in old age, and is often mistaken for it. There may be a previous history of manic-depressive illness, but a proportion of cases of mania in late life are first presentations, usually in association with underlying organic brain disease. Elderly manic patients are often exhausted and dehydrated, and so 'manic delirium' is a common presentation.

Other disorders

Anxiety states in elderly patients are unlikely to be mistaken for delirium, unless they are particularly severe. Similarly, paranoid states and schizophrenia rarely lead to diagnostic difficulty, although it should be noted that patients with these disorders are at an increased risk of developing delirium, either through self-neglect, or the effects of neuroleptic and anticholinergic medications. A number of other rare conditions in which cognitive, perceptual, affective, and behavioural disturbances occur, such as amnesic syndromes, epilepsia partialis continua, twilight states, the Charles Bonnet syndrome, neuroleptic malignant syndrome, and catatonia, may also resemble delirium. If the history and clinical examination are inconclusive, EEG may be helpful in making the diagnosis.

Epidemiology

The community prevalence of delirium increases with age, rising to 14 per cent in those aged 85 years and older. In medical and surgical inpatients, the rates of delirium vary considerably (prevalence, 10–30 per cent; incidence, 4–53 per cent), because of methodological and population differences. Similar rates are also found in studies of acute psychogeriatric admissions.[11] Some patient groups, such as those with hip fractures, have consistently higher rates. Other at-risk populations, such as nursing home residents, have received less systematic investigation, but the available evidence

suggests that they also have rates of delirium comparable to those found in elderly inpatients.

Aetiology

Almost any physical illness can give rise to delirium in elderly patients. The most common physical causes are listed in Table 8.5.1.1. In many cases the underlying cause is not obvious, and the delirium may be the most prominent presenting feature. The aetiology is commonly multi-factorial, and all contributory factors need to be identified and treated. As a rule, hyperactive delirium is more commonly due to infection and toxic/withdrawal states, whereas hypoactive delirium is more commonly due to metabolic abnormalities.

Table 8.5.1.1 Common causes of delirium in elderly patients

Drugs
Psychotropics
Hypnotics
Anticonvulsants
Anticholinergic drugs
Dopamine agonists
Analgesics
Anaesthetics
Alcohol withdrawal
Infection
Urinary tract infections
Pneumonia
Septicaemia
Ulcers, pressure sores, gangrene
Endocarditis
Postsurgical wound infection
Metabolic and endocrine
Electrolyte abnormalities
Uncontrolled diabetes
Hyper/hypothyroidism
Renal failure
Hepatic failure
Hypothermia
Malnutrition
Cardiovascular
Cardiac failure
Myocardial infarction
Vascular disease
Anaemia/polycythaemia
Respiratory
Pulmonary embolism
Pneumothorax
Pleural effusion
Intracranial
Trauma
Subdural haematoma
Stroke
Tumour
Epilepsy
Gastrointestinal
Perforation
Pancreatitis
Cholecystitis/cholangitis
Haemorrhage
Constipation

Drugs are an important cause of delirium in elderly patients, due to age-associated changes in their distribution, metabolism, and excretion. These pharmacokinetic changes are very variable, with the result that toxicity at apparently therapeutic doses is unpredictable. Certain drugs are particularly prone to cause delirium in elderly patients, for example those with anticholinergic activity. Tricyclic antidepressants, thioridazine, and benzhexol are particularly toxic in this respect, but many of the drugs commonly prescribed to elderly patients have some degree of anticholinergic activity, for example, digoxin, prednisolone, cimetidine, ampicillin, and warfarin. Individually, this activity may be small, but the cumulative effect can be significant if patients are on multiple medications.[12] Patients with Alzheimer's disease are particularly prone to develop delirium when given anticholinergic drugs, perhaps because their central cholinergic function is already impaired. In a minority of particularly vulnerable elderly patients, purely environmental and psychological insults are sufficient to cause delirium. The mechanisms of action in these cases are not known, but may involve factors such as sensory deprivation and stress responses via the hypothalamic-pituitary-adrenal axis.

Course and prognosis

Traditionally, delirium has been regarded as a transient condition that proceeds to either recovery or death. In the majority of cases, the delirium is brief, but about one-third of patients have prolonged or recurrent episodes.[13] Delirium is associated with increased short-term mortality in elderly patients, mainly because of the severity of the underlying illness. Delirium interferes with the processes of diagnosis, treatment, and rehabilitation, and as a result patients have longer hospital stays and higher rates of functional decline and discharge to nursing homes.[14] Increased length of stay and mortality are particularly associated with hypoactive delirium. In general, patients with hyperactive delirium appear to be less severely ill than those with hypoactive delirium; this may be due to differences in the cause of the delirium, or to the fact that hyperactive delirium is more likely to be identified and the causes treated.

Prospective studies have shown that the prognosis, in terms of persistent or recurrent symptoms, is relatively poor in elderly patients.[15] This is probably because those who experience delirium are a vulnerable group who are likely to develop the condition provided there is sufficient external insult. A proportion will also be suffering from a form of dementia, which will increase the vulnerability to delirium as it progresses. There is evidence that delirium is followed by persistent cognitive decline,[16] which raises the possibility that it (or the underlying cause) is a risk factor for the development or exacerbation of dementia.

Evaluation of treatment

Evidence regarding the efficacy of treatments for delirium is sparse (Chapter 4.1.1). The cholinergic hypothesis of delirium raises the possibility that cholinergic agonists, such as the cholinesterase inhibitors licensed for the treatment of Alzheimer's disease, may be of value in the prevention and treatment of delirium.

Management

There are four important steps in the management of delirium[17]:

Address the underlying causes (see above)

Behavioural control

This aspect of delirium management can be divided into pharmacological and non-pharmacological strategies. Non-pharmacological interventions in delirium are aimed at reducing the confusing, frightening, and disorienting aspects of the hospital environment in which most patients find themselves. They have received little formal evaluation, but features such as good lighting, low noise levels, a visible clock, and the reassuring presence of personal possessions and familiar individuals, such as relatives, are thought to be helpful. Any invasive intervention, including personal care tasks, should be introduced and explained simply, slowly, clearly, and repeatedly before it is carried out. Holding the patient's hand while talking helps to focus attention and provides reassurance.

The drug treatment of the symptoms and behaviours of delirium in the elderly is similar to that of younger patients, although it is necessary to start with lower doses, such as haloperidol 0.5 to 2 mg orally, or intramuscularly if necessary, repeated until the disturbance is controlled. Prescriptions should be for short periods only (up to 24 h) to encourage review of the effects and the necessary dosage. Once the delirium has resolved, the medication should be reduced/discontinued over a period of 3 to 5 days. If the patient cannot tolerate typical or atypical neuroleptic drugs, then a benzodiazepine (for instance, diazepam, lorazepam, or alprazolam) should be used instead.

Prevent/treat complications

The complications that befall patients with delirium probably contribute to the adverse outcomes associated with this condition. For example, hyperactive delirium is associated with falls during the hospital admission, whereas hypoactive delirium is associated with the development of pressure sores. Other complications of delirium include urinary incontinence, sleep disturbance, malnutrition, and immobilization; all of these problems should be anticipated and prevented where possible.

Rehabilitation and family support

Given the risk of functional decline following delirium, every effort should be made to return the patient to their pre-morbid level of functioning. ADL capacity should be assessed regularly, and independence encouraged where possible. The patient's family need to be involved in the rehabilitation process, as they will be largely responsible for aftercare following discharge. They should know that delirium is often recurrent, and be advised about the early signs of this. Indeed, delirium is a useful marker of vulnerability, and of the need for more intensive community aftercare.

Prevention

The modern hospital environment contributes significantly to the development of delirium in elderly patients, and multi-component interventions to improve poor clinical practice have been shown to reduce cost-effectively the incidence of delirium in elderly inpatients.[18] The following areas are important:

- *Prescribing*

Avoid where possible any drugs with known deliriogenic potential, particularly in at-risk individuals such as those with Alzheimer's disease. There should be regular review of the drug chart, with the aim of keeping the number of drugs to the minimum necessary. Non-pharmacological sleep-promotion strategies should be used in preference to hypnotic drugs.

- *Ward environment and routines*

These should aim to minimize disorientation, sensory impairment, and sleep deprivation. Patient mobility should be encouraged, as should adequate food and fluid intake. Medical and nursing staff should be trained to recognize and manage delirium.

- *Surgical routines*

Good preoperative, perioperative, and postoperative care (especially with regard to infection control, blood pressure, and oxygenation) will reduce the risk of postoperative delirium.

Further information

American Psychiatric Association. (1999). *Practice guideline for the treatment of patients with delirium*. American Psychiatric Association, Washington, DC.

British Geriatrics Society. (2006). *Guidelines for the prevention, diagnosis and management of delirium in older people in hospital*. British Geriatrics Society, London.

Byrne, E.J. (1994). *Confusional states in older people*. Edward Arnold, London.

Lindesay, J., Rockwood, K., and Macdonald, A. (2002). *Delirium in old age*. Oxford University Press, Oxford.

Lipowski, Z.J. (1990). *Delirium: acute confusional states*. Oxford University Press, New York.

References

1. Bowler, C., Boyle, A., Branford, M., *et al.* (1994). Detection of psychiatric disorders in elderly medical in-patients. *Age and Ageing*, **23**, 307–11.
2. Treloar, A.J. and Macdonald, A.J.D. (1997). Outcome of delirium: Parts 1 and 2. *International Journal of Geriatric Psychiatry*, **12**, 609–18.
3. Liptzin, B., Levkoff, S.E., Cleary, P.D., *et al.* (1991). An empirical study of diagnostic criteria for delirium. *The American Journal of Psychiatry*, **148**, 451–7.
4. Levkoff, S.E., Liptzin, B., Cleary, P.D., *et al.* (1996). Subsyndromal delirium. *The American Journal of Geriatric Psychiatry*, **4**, 320–9.
5. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). Mini-mental state-a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **2**, 1891–8.
6. Trzepacz, P.T., Baker, R.W., and Greenhouse, J. (1988). A symptom rating scale for delirium. *Psychiatry Research*, **23**, 89–97.
7. Inouye, S.K., van Dycke, C.H., Alessi, C.A., *et al.* (1990). Clarifying confusion: the confusion assessment method. *Annals of Internal Medicine*, **113**, 941–8.
8. Albert, M.S., Levkoff, S.E., Reilly, C., *et al.* (1992). The delirium symptom interview: an interview for the detection of delirium symptoms in hospitalized patients. *Journal of Geriatric Psychiatry and Neurology*, **5**, 14–21.
9. Inouye, S.K., Viscoli, C.M., Horowitz, R.I., *et al.* (1993). A predictive model for delirium in hospitalised elderly medical patients based on admission characteristics. *Annals of Internal Medicine*, **119**, 474–81.

10. Inouye, S.K. and Charpentier, P.A. (1996). Precipitating factors for delirium in hospitalized elderly persons. Predictive model and interrelationship with baseline vulnerability. *The Journal of the American Medical Association*, **275**, 852–7.
11. Lindesay, J., Rockwood, K., and Rolfson, D. (2002). The epidemiology of delirium. In *Delirium in old age* (eds. J. Lindesay, K. Rockwood, and A. Macdonald), pp. 27–50. Oxford University Press, Oxford.
12. Tune, L., Carr, S., Hoag, E., *et al.* (1992). Anticholinergic effects of drugs commonly prescribed for the elderly: potential means of assessing risk of delirium. *The American Journal of Psychiatry*, **149**, 1393–4.
13. Rudberg, M.A., Pompei, P., Foreman, M.D., *et al.*(1997). The natural history of delirium in older hospitalized patients: a syndrome of heterogeneity. *Age and Ageing*, **26**, 169–74.
14. Inouye, S.K., Rushing, J.T., Foreman, M.D., *et al.* (1998). Does delirium contribute to poor hospital outcome? A three-site epidemiologic study. *Journal of General Internal Medicine*, **13**, 234–42.
15. Levkoff, S., Evans, D., Liptzin, B., *et al.* (1992). Delirium, the occurrence and persistence of symptoms among elderly hospitalised patients. *Archives of Internal Medicine*, **152**, 334–40.
16. Jackson, J.C., Gordon, S.M., Hart, R.P., *et al.* (2004). The association between delirium and cognitive decline: a review of the empirical literature. *Neuropsychological Review*, **14**, 87–98.
17. Marcantonio, E. (2002). The management of delirium. In *Delirium in old age* (eds. J. Lindesay, K. Rockwood, and A. Macdonald), pp. 123–51. Oxford University Press, Oxford.
18. Inouye, S.K., Bogardus, S.T., Charpentier, P.A., *et al.* (1999). A multicomponent intervention to prevent delirium in hospitalized older patients. *The New England Journal of Medicine*, **340**, 669–76.

8.5.1.1 Mild cognitive impairment

Claudia Jacova and Howard H. Feldman

Introduction

Within the cognitive functioning continuum from normal ageing to dementia three broad states can be distinguished: normal functioning for age, clear-cut impairment meeting diagnostic criteria for dementia, and mild cognitive impairment (MCI), which falls below normal but short of dementia in severity (Fig. 8.5.1.1.1). There is active debate over what MCI is, how to define and classify this state, and where to set its borders on the described continuum.[1] Some definitions depict MCI as the tail-end of normal cognitive ageing whereas in other definitions MCI embodies the early clinical manifestation of Alzheimer Disease (AD) and other dementias. In 2003, the key elements of different MCI definitions were integrated into a consensus diagnostic and classification framework,[2] thus establishing some common ground in a field that is still evolving. MCI has also been positioned as a potentially important target for early treatment interventions to delay progression to dementia.

Nosologically, MCI is not currently included as a diagnostic entity in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR)[3] and the *International Classification of Diseases*, 10th revision.[4] The diagnostic categories of *Mild Neurocognitive Disorder* (DSM-IV-TR) and *Mild Cognitive Disorder* (ICD-10) are similar to MCI because they require the presence of cognitive impairment but these categories can only be assigned if a specific neurological or general medical condition can be identified to account for the cognitive symptoms. Much of the current condition

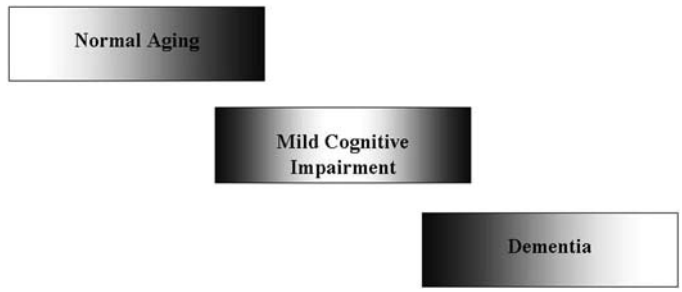


Fig. 8.5.1.1.1 Theoretical continuum from normal ageing to dementia: darker shading indicates areas of overlap between adjacent states and increased diagnostic challenge. (Reproduced from R. Petersen *et al.* Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals, *The Journal of the American Medical Association*, **273**, 1274–78, copyright 1995, The American Medical Association.)

of MCI does not fit as it has no aetiologic specification. Nevertheless, MCI is increasingly a presenting condition in primary and specialized settings of care. Medical practice guidelines have recognized MCI as a risk state for dementia and recommend careful clinical evaluation and monitoring of individuals with this diagnosis.[5,6]

Nosology

The current nosological entities within the general MCI framework include a variety of definitions and capture overlapping but not identical conditions in the ageing population (Fig. 8.5.1.1.2).

Age-associated memory impairment (AAMI)

AAMI describes healthy individuals over the age of 50 that experience memory decline. Formal diagnostic criteria require complaints of memory loss, performance on objective memory tests falling at least 1 SD below norms for young individuals, and intellectual functioning normal. AAMI cannot be diagnosed if there is a

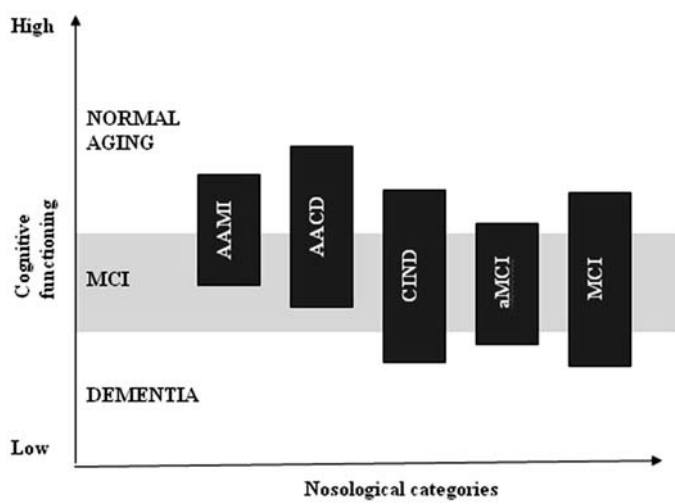


Fig. 8.5.1.1.2 MCI nosological entities on the continuum from normal ageing to dementia. (Reproduced from H.H. Feldman and C. Jacova, Mild cognitive impairment, *The American Journal of Geriatric Psychiatry*, **13**(8), 645–55, copyright 2005, American Association of Geriatric Psychiatry, Lippincott Williams & Wilkins.)

neurological, psychiatric, or medical condition that can account for the impairment.[7]

Age-associated cognitive decline (AACD)

The AACD category covers impairments in any domain affected by ageing, including learning, memory, attention, thinking, language, and visuospatial function. There must be self- or informant-reported cognitive decline over at least 6 months and performance on objective cognitive tests at least 1 SD below age- and education-appropriate norms. The cognitive impairment cannot fulfil dementia criteria and is not accounted for by systemic, neurological, or psychiatric disorders.

Cognitive impairment not dementia (CIND)

CIND includes all individuals that cannot be classified as cognitively normal or as demented. CIND has been applied both in population- and clinic-based studies.[8,9] This diagnostic label is assigned by clinical judgement when there is memory and/or cognitive impairment insufficient to meet DSM criteria for dementia, without exclusions related to underlying aetiologies. There are to date no operational criteria for this category. Because of its inclusiveness CIND encompasses a range of aetiologies that must be disentangled to be clinically meaningful.[8,9]

Amnestic mild cognitive impairment (aMCI)

This amnestic condition is defined as a clinical disorder that describes a transitional state between normal ageing and AD. It is characterized by memory impairment in the context of otherwise preserved abilities. The diagnostic criteria for aMCI require memory complaint preferably corroborated by an informant, objective memory impairment for age and education, largely normal general cognition, essentially intact activities of daily living (ADLs), and the absence of dementia. Objective memory impairment, though not anchored to a specific cut point, is generally ≥1.5 SD below appropriate norms.[10]

Mild cognitive impairment (MCI): international working group criteria

The MCI concept has recently been broadened to encompass multiple patterns of cognitive impairment including amnestic, non-amnestic, single- or multiple-domain deficits.[2] In this framework the classification of MCI requires multiple steps. First, individuals should be judged as neither normal nor demented. Second, there should be evidence of cognitive decline, supported by self and/or informant reports, impairment on objective cognitive tests, or evidence of decline over time on these tests. Third, activities of daily living should be mainly preserved, with the provision that complex ADLs can be minimally impaired.[2] Like CIND, this MCI category recognizes multiple aetiologies underlying impairment, and requires their identification.

Clinical staging scales

Studies of MCI frequently utilize clinical staging scales both to define the inclusion criteria as well as to track outcomes. The Clinical Dementia Rating (CDR) scale[11] distinguishes five stages of dementia severity, with a stage of questionable dementia (CDR 0.5), between the stages of healthy (CDR 0) and mild dementia (CDR 1). CDR 0.5 is most often applied to MCI; however, this stage also can include those with functional impairment who meet dementia criteria. Similarly, the Global Deterioration Scale (GDS),[12] which distinguishes seven stages of impairment, overlaps with MCI at stage 2 (normal with a subjective complaint) or stage 3 (subtle deficits in cognition and occupational/social activities), whereas individuals with mild dementia may receive a GDS stage 3 or 4. The mapping of MCI onto these staging scales has not yet been fully reconciled.

Epidemiology

Prevalence

Prevalence estimates for MCI will naturally vary according to the definition, to age and to the setting. In population-based studies, AAMI has been estimated to affect up to 38.4 per cent, and AACD between 21 and 35.2 per cent, of individuals aged 60 or older.[13] CIND has been reported to affect between 16.8 and 23.4 per cent of individuals aged 65 or older. The prevalence of aMCI has been much lower, at 3 to 6 per cent in similar age groups.[13] Within the broad MCI classification, the multiple domains and single non-memory domain subtypes have been described as roughly twice as frequent as the amnestic subtype, with multiple domain impairment estimated to affect 16 per cent of individuals.[14,15] The prevalence of AACD, aMCI, MCI, and CIND varies with age, with a two- to threefold increase from age 65–74 to >85 in CIND (Fig. 8.5.1.1.3).[8,16] The prevalence of MCI and related conditions within the referral clinic setting is much higher than population estimates.[17]

Natural history

(a) Progression rates to dementia

While the rate of progression to dementia is 1 to 2 per cent per year for cognitively normal individuals aged 65 or older, the rates for all MCI entities are systematically higher and quite variable (Fig. 8.5.1.1.4). Whereas AAMI has low progression rates (1 to 3 per cent per year) and is closest to a normal population, for all other categories most studies report rates between 10 and 15 per cent per

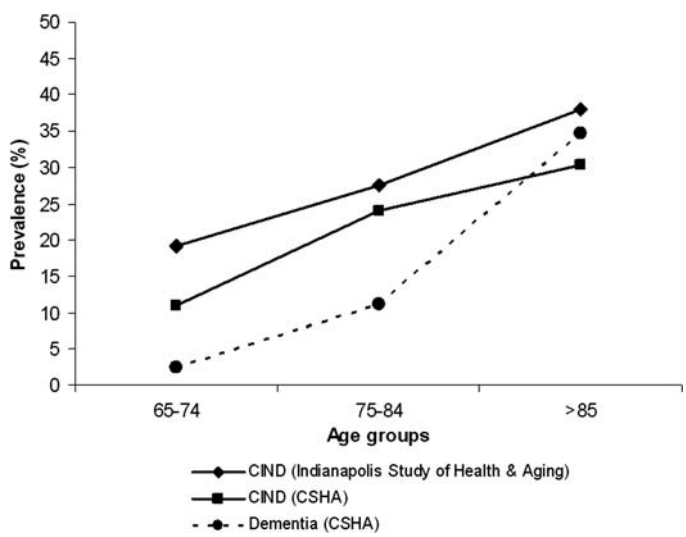


Fig. 8.5.1.1.3 The prevalence of CIND with increasing age reported in the Canadian study of health and ageing (CSHA)[8] and the Indianapolis study of health and ageing.[16]