EDITORIALS

Evolution of outcome measures in schizophrenia
Tom Burns

REVIEW ARTICLES

Symptom rating scales and outcome in schizophrenia
Ann M. Mortimer

Social outcomes in schizophrenia
Stefan Priebe

Patient-reported outcomes in schizophrenia
Rosemarie McCabe, Marya Saidi, and Stefan Priebe

Outcomes of public concern in schizophrenia
Iain Kooyman, Kimberlie Dean, Samuel Harvey, and Elizabeth Walsh

Hospitalisation as an outcome measure in schizophrenia
Tom Burns

Health economic measures in schizophrenia research
Paul McCrone

Cognition as an outcome measure in schizophrenia
Michael S. Kraus and Richard S. E. Keefe

Neuroimaging and other neurobiological indices in schizophrenia: relationship to measurement of functional outcome
John L. Waddington

Outcome measures in early psychosis: Relevance of duration of untreated psychosis
Swaran P. Singh
Adverse effects of antipsychotics as outcome measures
Samantha Hamer and Peter M. Haddad
Br J Psychiatry 2007 191: s64-s70.

Schizophrenia outcome measures in the wider international community
Mohan Isaac, Prabhat Chand, and Pratima Murthy

Measurement of long-term outcomes in observational and randomised controlled trials
Richard Hodgson, Chris Bushe, and Robert Hunter
Br J Psychiatry 2007 191: s78-s84.
Evolution of outcome measures in schizophrenia

TOM BURNS

Summary  Outcome has been a key concern in schizophrenia since Kraepelin first identified the disorder. The outcomes that researchers and clinicians use and value in their work with it have changed to reflect evolving theories about it and the available interventions. This supplement tracks those changes and examines the merits and use of structured approaches to understand this most complex disorder.

Declaration of interest  T.B. has received payments for lectures and consultancies from Eli Lilly, Janssen and Otsuka in the past 5 years.

Schizophrenia has been with us as an identified illness for over a century. Kraepelin distinguished it as ‘dementia praecox’ in 1896 (Kraepelin, 1919) separating it from the broad spectrum of psychoses seen within his clinic, and Bleuler renamed it schizophrenia in 1908 (Bleuler, 1950). Illnesses are usually identified and defined in terms of their clinical presentation, course and outcome. Kraepelin’s identification of what we now call schizophrenia rested almost exclusively on course and outcome. He conducted his research in Dorpat, in what is now Estonia, where he had taken a post because his recent marriage was incompatible financially or practically with his research post. German was the professional language in Dorpat but was not spoken by Kraepelin’s patients which precluded detailed clinical interviews, and hence his focus on the pattern of illnesses.

RECOVERY

Kraepelin’s major distinction was between dementia praecox and manic–depressive (bipolar) disorder in which there were periods of substantial recovery and even discharge. He formed a very pessimistic view of the outcome in schizophrenia, and was convinced that recovery was very rare, or even impossible, and deterioration almost inevitable.

One consequence of Kraepelin’s view of the long-term outcome in schizophrenia has been the persistence in psychiatric textbooks and teaching of an excessively gloomy estimate of the outcome (van Os et al, 2006). Kraepelin’s perspective, like that of many current clinical psychiatrists, was shaped by his institutional experience. He spent the bulk of his professional life working with the patients who did not recover, or those who only partly recovered, and was ignorant of those who got well and moved on. In contrast, Bleuler spent long periods in conversation with patients (including discharged and recovered patients) exploring their experiences, and took a less gloomy view of the disorder.

This supplement explores the domains of outcome measurement as they have been used in schizophrenia research. It would be misleading to suggest that the evolution of these different approaches demonstrates a single, unified development. However there is something of an evolution that can be discerned which I will attempt to outline.

LONG-TERM FOLLOW-UP STUDIES

Both of Kraepelin’s main proposals about schizophrenia have been subject to extensive revision. The clear distinction between schizophrenia and bipolar disorder that he introduced has been challenged (Moller, 2003), as indeed has the very coherence of the concept of schizophrenia as a disorder (Bentall & Beck, 2004). Similarly his gloomy appraisal of the outcome has been challenged, most convincingly by careful follow-up studies. This supplement contains a series of papers which explore the sophisticated range of outcome measures (and to a lesser extent the investigatory techniques) that have come into use to explore and compare outcomes in this most complex of disorders.

Randomised controlled trials v. naturalistic studies

Hodgson et al (2007), this supplement review the use of longer follow-up studies to determine the outcome of schizophrenia. In the past three decades randomised controlled trials (RCTs), and particularly RCTs of antipsychotic medication, have come to dominate research in schizophrenia. Most of these are short-term (many ultra-short-term), often only weeks, and have limited follow-up (rarely beyond a year).

The classical follow-up studies spanning years and decades (Harding et al, 1987; Ciompi, 1988) confirmed the reality of recovery in a substantial proportion of people with schizophrenia. These very long-term outcomes are essentially a description of the natural history of the disorder rather than a response to any specific intervention. Not surprisingly they have been viewed as less relevant to the practising clinician.

Hodgson et al consider the newer generation of long-term studies – usually of a year or so. Many of these are RCTs rather than naturalistic observational studies and often focus on drop-out from treatment or changes in treatment as proxies for clinical response (Lieberman et al, 2005). One reason for favouring RCTs in schizophrenia is the belief that quasi-experimental studies overestimate treatment effects. However, a series of exchanges in the New England Journal of Medicine questioned this and demonstrated that effect sizes in experimental and quasi-experimental studies were remarkably similar (Concato et al, 2000). There has been something of a resurgence of cohort studies in actively treated schizophrenia, such as the European Schizophrenia Outpatient Health Outcomes (SOHO) study (Haro et al, 2003) and the Schizophrenia Care and Assessment Programme (SCAP) study (Burns et al, 2006). These often use convenient or consecutive sampling and pragmatic, simple, clinician-rated outcome measures.

Hodgson et al point to the increasing call for more naturalistic, long-term treatment data from regulatory organisations such as the National Institute for Health and Clinical Excellence (NICE) in the UK. Despite the advantages of these studies in terms of sample size and generalisability,
they raise important methodological questions concerning design and analysis. Hodgson et al also highlight the neglected potential of post-marketing surveillance conducted by pharmaceutical companies in illuminating long-term outcomes in the era of active management.

**SYMPTOM OUTCOMES**

Bleuler (1950) and Jaspers (1963) explored the form of schizophrenic experiences as the essential route to understanding the disorder. Both emphasised the importance of the structure, or form, of pathological experiences rather than their content, and this diagnostic approach was codified in 1923 by Kurt Schneider in his ‘first-rank’ symptoms (Schneider, 1959). Schneider was concerned to improve diagnostic reliability, in particular from individual interviews, in contrast to the extended familiarity with the patient practised by Bleuler. It is unlikely that Schneider rejected Bleuler’s understanding of the basic pathology of schizophrenia, rather that he thought that hallucinations, thought disorder and delusions were more likely to be identified and recorded reliably. This ‘Schneiderian’ approach lends itself to modern treatment trials. It is these productive symptoms that respond rapidly to successful treatment, are more obvious and more easily quantifiable.

**Structured symptom scales**

Mortimer (2007, this supplement) outlines the early pre-eminence of using these positive symptoms to measure outcome in schizophrenia treatment trials. Initially, composite instruments such as the Present State Examination (PSE; Wing et al, 1974), which was designed more for diagnosis than outcome measurement, were used but these were soon superseded by rating scales specifically designed to measure symptom change. Mortimer presents the three rating scales that have been most extensively used in schizophrenia trials: the Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962); the Positive and Negative Syndrome Scale (PANSS; Kay et al, 1987), which was developed from the BPRS; and the single rating of Clinical Global Impression (CGI; American Psychiatric Association, 1994). Although the primary use for each of these three involves the computation of a single sum-score, the first two have sub-scales whose analysis has helped to refine and to understand possible clinical sub-categories of schizophrenia.

Mortimer points to the primacy of symptom-rated outcomes in schizophrenia. Although no longer the only outcome measures, it is often argued (perhaps with the exception of studies of cognitive-behavioural therapy) that these other outcomes (e.g. social, vocational, well-being, satisfaction) derive from successful control of the illness and do not occur without it. Symptomatic change is proposed as the key response to treatment – whether pharmacological or other.

**COGNITIVE AND NEUROBIOLOGICAL OUTCOMES**

Psychopathology was for several decades the medium for trying to understand what schizophrenia was and its underlying processes. However, classical psychopathology cannot explain all the variation in response to treatment, nor indeed the persisting social disabilities. With increased sophistication in experimental and clinical psychology, the role of cognitive functioning in the natural history of schizophrenia has become a focus of intense investigation.

**Cognitive function as a core feature**

Kraus & Keefe (2007, this supplement) review current understanding of cognitive functioning in schizophrenia. They point out that problems in cognitive functioning are core to the disorder (not simply consequences of either symptoms or treatments) and they remind us that both Kraepelin and Bleuler considered cognitive decline as an inherent feature of the condition. After a long period in the shadows of Schneiderian symptomatology, cognitive dysfunction has returned centre-stage with the increasing recognition (Green, 1996) that it is responsible for much social impairment. Kraus & Keefe overview a wide battery of available tests but there is clearly still some way to go before cognitive testing is used either in routine clinical practice or in most treatment studies. They predict that testing of cognitive function will eventually not only become a part of routine outcome assessment but will probably fundamentally reshape how we conceive the disorder.

**Neurobiology**

Waddington (2007, this supplement) suggests that we may soon begin to realise the potential of a wide range of neurobiological indices and measurements in schizophrenia (the various forms of imaging, neurodevelopmental indices, genomics, proteomics, metabolomics and apoptotic mechanisms). He reviews the use of these techniques to explore the pathobiology of schizophrenia and their increasing potential to chart outcome. Apart from some first-episode psychosis samples most studies to date have been cross-sectional rather than longitudinal. Although these techniques are not yet useable in prediction at the individual patient level they may soon generate useable ‘biomarkers’.

**PATIENT-REPORTED OUTCOMES**

Despite the future potential of cognitive testing and neurobiological measures, the diagnosis and monitoring of progress in schizophrenia takes place mainly in face-to-face interviews between patient and clinician. All the outcomes mentioned so far (and, indeed, most of those to come) are rated by the clinician or researcher. However, mental illnesses are characterised predominantly by the patient’s private experiences, and clinicians rely on patients recounting to make ratings. In sharp contrast to the situation in depression and anxiety there has been a much more limited development of ratings to be completed by people with schizophrenia patients themselves. McCabe et al (2007, this supplement) review such patient-reported outcomes. They explore why there has been so little work in the field, suggesting that clinicians might mistrust patients’ judgments because of the severity of their disorder, and why this reluctance to use patient self-reports is being overcome. Many of the same concerns have been raised in the field of social outcomes which are reviewed by Priebe et al (2007, this supplement). Despite this, patient-reported outcomes have been shown to predict service use and the increasing ‘consumer’ voice in mental health.

**Consumerism and well-being**

Patients and their families want their appraisal of the situation to be taken more seriously. They want to say what should be judged in outcomes and to have some control over their estimation. One consequence of this is the broadening of the measures used in outcomes – satisfaction...
with services, met and unmet needs, therapeutic relationships and a greater emphasis on 'real world' outcomes such as jobs and accommodation. There has also been an increased focus on assessment of personal well-being or function, with the development of scales for empowerment, self-esteem, sense of coherence and recovery. These outcomes are patient- or person-centred rather than disease- or disorder-centred.

**Use of patient-reported outcomes**

McCabe *et al* warn against the simultaneous use of too many such scales in schizophrenia research. They point to their high level of covariance, which may indicate a single common underlying factor (probably dependent on mood). Consequently patients who report positively on therapeutic relationship are likely also to report high satisfaction with services and empowerment, etc. However, there are differences between the scales and good evidence for validity and reliability for many.

One potential development explored here is the routine use of outcome measure and McCabe *et al* compare the effectiveness of such scales in feedback during the clinical interview (their own work) or delayed feedback using questionnaires. A number of exciting possibilities are opening up in this area to improve communication and understanding in the clinical interview and to measure progress and outcome.

**CONTEXT-DEPENDENT OUTCOME MEASURES**

Outcome measurement is used clinically mainly to judge improvement in an individual patient and in research studies to judge the efficacy of a specific treatment. The rater-recorded assessments covered so far (longitudinal outcomes, symptoms, cognitive and neurobiological measures), and to some extent the patient-reported outcomes, also serve an important function in exploring the nature of the disorder itself.

These outcomes are independent of treatment and of the local social context or the system of care. However, schizophrenia is a disorder with a profound impact on the wider society and one that absorbs a major proportion of healthcare resources. A range of outcome measures have been developed which go beyond the progress of the disease in an individual or group of individuals and measure either the impact of social and clinical responses to the disease or the wider impact of the disease on society.

**ADVERSE DRUG EFFECTS**

Treatments are not risk free. As Hammer & Haddad (2007, this supplement) point out antipsychotic drugs are associated with a wide range of adverse effects which can lead to distress and impaired quality of life. The psychological price that patients may pay for a reduced risk of relapse, or for a prompt reduction in acute symptoms, can include tiredness, sluggish thinking, or even frank depression. Added to this are physical problems such as stiffness, akathisia, reduced sexual functioning and a whole range of longer-term hormonal and weight problems. These are a significant burden on individuals who already have schizophrenia to contend with. These adverse effects can also lead to poor adherence to medication if patients do not consider the benefits of the medication to outweigh the adverse effects. Some of these side-effects are also obvious to others (e.g. stiffness, tremor, weight gain) and can contribute to stigma.

Hamman & Haddad outline the various neglected sources of data on adverse effects and the inconsistency of their recording. They argue for greater attention to these outcomes (and patients certainly do consider them outcomes) and for the need to obtain information from all possible sources rather than stick to a single protocol. Like Hodgson *et al* they believe that total discontinuation rates for antipsychotics (for whatever reason – poor clinical response or side-effects, or poor clinical response and side-effects) are an increasingly practical and valid outcome measure in schizophrenia management. They argue for the recording of adverse effects in clinically, rather than statistically, meaningful ways (e.g. the number of patients who became obese during a trial rather than the mean weight gain, the number of patients who developed significant akathisia rather than the mean increase in akathisia score). They propose that more careful and routine measurement of side-effects should lead to greater openness in discussing their likelihood with patients.

**WIDER SOCIETAL OUTCOMES**

Adverse drug effects, although dependent on local treatment regimes, are outcomes that are of direct relevance to the treated patient. There is, however, a range of outcomes that might be of equal or greater interest to others (the healthcare system, society at large) than to the patient. Kooyman *et al* (2007, this supplement) consider outcomes of public concern in schizophrenia. The care of people with schizophrenia is profoundly affected by infrequent but high-profile consequences of their illness. It is these dramatic and public outcomes such as violence or suicide, or socially unacceptable outcomes such as homelessness and vagrancy, that set the policy agenda and attract or deflect investment in mental healthcare.

Kooyman *et al* outline the major areas – violence, victimisation, suicide and self-harm, substance misuse, homelessness and unemployment. For each they present what is known about the major risk factors and the methodological problems presented in attempting to measure these outcomes. Unlike patient-centred outcome measures these are mainly direct measures which do not need extensive testing of validity, sensitivity and reliability as do scales developed from psychological theory. However, the reliability and comparability of different ways of collecting these data is an equally difficult challenge – for example there is no consensus on the timescales of recording (e.g. past month, or week, or lifetime) across different statistics.

Employment or homelessness can be sensitive indicators of the adequacy of local services, and episodes of violence or suicide (albeit individually rare) can alert observers to failing services. Even if it can be argued that many of these lie outside the power of psychiatry to influence, failing to take account of them will certainly lead to criticisms of services and demands for changes.

**HOSPITALISATION**

Hospitalisation is one of the most common outcome measures used in mental health services research, particularly in RCTs (Catty *et al*, 2002). The strength of this outcome, like many of the wider societal outcomes, is that it is a 'hard' outcome. Although there may be some difficulties in collecting the data consistently, it is clear what it means and it does not require much interpretation. Later in this supplement (Burns, 2007) I explore the differing ways in which hospitalisation has been used in mental health services research for.
psychoses – from simply ‘admitted/not admitted’, through number of admissions in a set follow-up period, to days in hospital and survival curves. In high-income countries, hospitalisation is a fairly good proxy for relapse in schizophrenia (although as Isaac et al (2007, this supplement) point out this is not so in low- and middle-income countries). The threshold will differ in different countries and healthcare systems and will depend on the quality of community care, but within a single study the difference will give a good indication of treatment effectiveness.

The problem with hospitalisation is, of course, its extreme context specificity, and I explore the risks inherent in generalising across different healthcare systems. Hospitalisation also lacks acceptability to an increasingly sceptical and critical consumer movement, as it appears to be an outcome relevant only to the service provider.

**DURATION OF UNTREATED PSYCHOsis**

Early intervention in psychoses is a major concern worldwide (Edwards et al, 2000). Services are being widely established to promote earlier detection and treatment of schizophrenia and other psychoses. The rationale is not just humanitarian (the reduction in the duration of untreated distress) but a growing recognition that the duration of untreated psychosis may have a major impact both on immediate recovery rates (Marshall et al, 2005) and on long-term outcomes and disability (Larsen et al, 2006). Singh (2007, this supplement) reviews this literature and explores whether or not the duration of untreated psychosis can be sensibly used as an outcome measure in its own right.

An independent effect of the duration of untreated psychosis on outcomes has been attributed to a direct ‘neurotoxic’ effect (Larsen et al, 2006) and to a ‘critical period’ in personal development when people may miss out on vital social development and may acquiring disabilities and patterns of behaviour with long-term consequences (Birchwood et al, 1998). Early intervention teams aim to reduce the duration of untreated psychosis and many have established extensive programmes of public education and outreach to achieve this.

Singh outlines the methodological problems in identifying when psychoses begin and when prodromal phases end. He questions the representativeness of the long durations of untreated psychosis reported in earlier studies and some of the very extensive reductions reported. However, the association between duration without treatment and outcome does give some support for its use as a service-level outcome measure in schizophrenia. A less clear picture of this association has been reported from India (see Isaac et al).

**ECONOMIC OUTCOMES**

Early intervention services may alter help-seeking patterns and reduce subsequent reliance on in-patient care. If this proves to be the case, then early intervention promises a substantial saving in healthcare costs. McCrone’s paper (2007, this supplement) on economic outcome measures in schizophrenia highlights how in-patient care accounts for a disproportionate amount of healthcare costs over the long term. Although most schizophrenia care is in the community, with brief in-patient care for acute relapses, in-patient care still accounts for most of the costs in high-income countries (although not necessarily in low- and middle-income countries – see Isaac et al). Indeed, the prominence of economic analyses in mental healthcare arose in part from the recognition that major savings could be achieved by modest shifts in the use of in-patient stays (Weisbrod et al, 1980).

McCrone outlines the range of economic outcome analyses that can be used to link costs with outcomes in schizophrenia care (cost minimisation, cost-effectiveness, cost consequence, cost-utility and cost–benefit analyses). The cost-effectiveness/cost-utility plane illustrates how judgements can be made about whether new interventions are economically indicated. However, where a more expensive approach produces a better result further consideration is needed.

Cost-utility analyses are developed to address this question using a standard outcome measure and calculating the cost of achieving one unit improvement with the treatments studied. The EuroQol (Williams, 1995) is probably the most widely used such unit of measurement but there are few schizophrenia outcome studies using it. Quality-adjusted life-years (QALYs) have the advantage that they can be used to compare cost-utility across different areas of health and are increasingly sought by health regulatory bodies such as NICE. However, McCrone points out their lack of sensitivity in schizophrenia and current work on incremental cost-effectiveness ratios which compare the cost of achieving an agreed improvement in a chosen outcome measure (e.g. the BPRS in schizophrenia). There is still some way to go with this approach, but work is ongoing on defining ‘clinically meaningful minimal changes’ in terms of symptom scores, and these should help to consolidate the use of incremental cost-effectiveness ratios in economic analyses.

Economic studies of schizophrenia care highlight to politicians and policy makers just how costly the disorder is to society. Schizophrenia is unlikely to compete successfully for public attention against cancer or cardiovascular disorders, but economic evaluations demonstrate clearly that it needs to be a top priority for service improvement and treatment research.

**INTERNATIONAL OUTCOMES**

The evolution of outcome measures outlined in this introduction and in this supplement appears to follow a coherent and logical pattern. However, this progression – from natural history, via symptoms to more scientific ratings as the technology becomes available, alongside more subjective measures and measures of quality of life as treatments improve, ending with specific measures of schizophrenia in consistent and highly evolved services – is only one path. Isaac et al (2007, this supplement) point out that the evolution of outcome measures in low- and middle-income countries has taken a different path, and not just because of a lack of resources. The pattern of the disorder is clearly influenced by the social context (Leff et al, 1992), with differences in the severity of outcome but also differences in local priorities and values. Social functioning is less affected in low- and middle-income countries but family support and burden are much more important. Unlike the West, the economics of care are markedly different with no suggestion that outcomes such as hospitalisation have any value in comparing interventions.

Isaac et al highlight the need to adapt Western instruments for use in very different settings but also to develop local instruments. The balance is between developing instruments that are highly sensitive to local conditions with high face validity and the need to compare outcomes internationally. Their review also brings us back
to some of the vital outcomes currently overlooked in the West – the impact on marriage and the very high early death rate in schizophrenia.

CONCLUSIONS

This supplement focuses on the varied aspects of outcome measures in schizophrenia. It has been suggested that the outcomes which are recorded and underpin decision-making have shown a clear evolution. However, this ‘march of progress’ view should not be taken too seriously – it is simply one way of organising the wealth of research in the area. Undoubtedly the introduction of active treatments has shifted the focus from naturalistic long-term outcome studies and improved the measurement of current status. However, the more recent developments reflect more complex drivers.

There are the enormous strides in technologies of both treatment and investigation. Effective drugs and service developments such as community mental health teams and assertive community treatment, mean that the outcome differences that must be measured are more sensitive but also broader. Most of the ‘softer’ outcome measures such as quality of life, social functioning and personal well-being are only of relevance in situations were symptom control is relatively well achieved. Cognitive functioning and the neurobiological parameters of the disorder have only recently come into their own as the science of their measurement matures.

As systems of care become more consistent and predictable they themselves affect the outcomes to be measured. The ubiquity of maintenance antipsychotic treatment necessitates greater attention to adverse effects of drug treatment. Assessment of net needs and patient satisfaction have become important patient-centred measures of how services are functioning. The costs of how services are functioning. The costs of how met needs and patient satisfaction have been used as a decent proxy outcome measure of clinical hard outcomes such as employment or adequacy of accommodation, etc.

From the opposite direction (from society and healthcare funders) there is increased pressure for closer scrutiny of the broader impact of schizophrenia. The rapid growth in mental health economic outcome studies indicates increased sensitivity to the social burden of the disorder, as does the current attention paid to risk and the wider societal outcomes of violence, victimisation, suicide, etc.

However, this neat attempt at imposing order is just that – an attempt, as Isaac et al remind us. The outcomes we need to measure are not fixed. They will continue to change as society’s preoccupations change, as our measurement technologies change and as treatments improve. What is clear, however, is that keeping abreast of developments in schizophrenia outcomes is a challenge for clinicians and researchers alike.

REFERENCES


Symptom rating scales and outcome in schizophrenia

ANN M. MORTIMER

Background Symptom rating scales are now well established in schizophrenia research but their scores are not the same as outcome.

Aims To appraise the usefulness of symptom rating scales in evaluating the outcome of people with schizophrenia.

Method Literature on the use of the Brief Psychiatric Rating Scale (BPRS) the Positive and Negative Syndrome Scale (PANSS) and the Clinical Global Impression (CGI) in schizophrenia research was studied.

Results Scales were designed to make diagnoses, to categorise patients, syndromes or both, and to demonstrate antipsychotic efficacy, as well as to measure outcome. There is much redundancy both between and within scales. Early work suggests limited concurrent validity with external outcome variables. Data are at best ordinal and there are particular difficulties in equating outcome with percentage changes in scores. The concept of remission, which uses absolute item score thresholds with a duration criterion, is a promising outcome measure.

Conclusions Symptom rating scale scores can only comprise a limited part of outcome measurement. Standardised remission criteria may present advantages in outcome research.

Declaration of interest A.M.M. has received funding from several pharmaceutical companies.

Outcome measures are important in schizophrenia because we need to identify whether outcomes are modified by the medications and psychosocial interventions which we offer. Leaving aside social cultural and environmental factors, before the antipsychotic era it is unlikely that outcome was influenced by anything other than the intrinsic nature and severity of the schizophrenic illness. Providing basic nursing care and protection probably influenced negative outcomes to some extent.

Outcome is not a unitary construct defined simply by lack of symptoms: personal and social function, cognition and quality of life must be of substantial relevance. Other aspects such as economic outcome, although important to commissioners and providers of services, might be of limited consequence to clinicians and patients, who naturally focus on professional and consumer (satisfaction) viewpoints respectively. Hence, outcome evaluation applied to services differs from that applied to patients.

Symptom rating scales in schizophrenia were not initially designed to assess the efficacy of antipsychotic drug treatments. Nevertheless, they have been used in this role more than any other. This is not surprising as antipsychotic drugs are used primarily to control patients’ symptoms; the underlying neuroscience is consistent with this, and not with any direct therapeutic effects on cognition, personal and social function, or quality of life (unless mediated by symptom control). Although such distal effects have been proposed, there are numerous independent variables which influence these aspects of outcome (e.g. upbringing, premorbid personality and adjustment, intellect and mood, social circumstances and availability of a support network). Furthermore, it has been proposed that antipsychotic drugs, particularly conventional antipsychotics, have little effect on negative symptoms of schizophrenia. Negative symptoms are one of the most clinically important targets, and overlap with cognition and function (Mortimer & Spence, 2001).

SYMPTOM RATINGS AS OUTCOME MEASURES

Although there is evidence that changes in distinct psychopathological dimensions differentially influence broader aspects of outcome (Van Os et al, 1996) it is now accepted that fixed factors such as duration of untreated psychosis, gender, age of onset and family psychiatric history make a substantial contribution (Murray & Van Os, 1998). Symptom rating scales can be viewed as quantifying the skilled clinician’s judgement of current psychopathology, and change over time. The worth of routine use of such rating scales in ordinary clinical practice is the subject of continuing debate; the clinician makes an initial, comprehensive assessment of the patient, and reviews this as treatment proceeds and the final outcome becomes clearer. The added value of a highly structured approach can be questioned in a clinical review of an individual patient’s progress. Most patients manifest only a minority of the range of possible symptoms and generally do not develop too many new symptoms during treatment.

In routine practice, symptom scales are perhaps little more than a formalised guide to what the clinician should be doing already. They have specific utility in training junior staff in the full range of psychopathology they are likely to encounter, and the finer points of mental state examination. Repeated scores, represented graphically, may have some utility in communicating a patient’s progress to other clinicians. In research, symptom rating scales in schizophrenia will inform the investigator what is the nature and ‘volume’ of symptoms experienced by the patient, and the magnitude of any change over time.

Limitations Symptom rating scale data can never be anything more than ordinal; the overall total of symptom item scores will often lump together categorical data, containing symptoms associated in clusters, such as the positive, negative and disorganisation syndromes. Specific syndrome scores derived from scales may have more utility than the total score regarding an overall perspective. Current thinking includes that schizophrenia syndromes may comprise positive...
(disorganisation and reality distortion) and negative categories, with non-negative affective symptoms (mostly depressive) in a significant minority of patients. Consequently three or four syndrome scores in the context of a defined range may give a reasonable ‘snapshot’ of a patient’s current clinical status. Such quantification may inform judgement regarding aetiology, treatment and prognosis (Van Os et al, 1996). For example, negative symptoms are known to have adverse consequences for personal and social function and cognition (Rocca et al, 2005). By contrast, even extensive, but isolated, reality distortion may generate minimal functional consequence, whereas disorganisation syndrome is usually very disruptive (Schulberg et al, 1999). Depression may arise from several sources, with varying outcome (Emstley et al, 1999). Such data have implications for treatment interventions. The Clinical Global Impression-Schizophrenia scale (CGI-SCH; Haro et al, 2003) represents, conceivably, a step in this direction although its positive, negative, depression and cognitive scores are rated according to judgement of severity rather than from items comprising these syndromes.

The value of symptom item or even syndrome score totals per se is increasingly questioned in the determination of outcome status. A more patient-centred definition of outcome, stressing personal and social function, is often viewed as more practical than the presence or absence of esoteric phenomena (symptoms), which may have little bearing on subjective experience or uptake of healthcare. Influential work has attempted to explore the meaning and consequences of delusions and hallucinations for patients (Chadwick & Birchwood, 1995), but scales derived from this work are not in widespread use outside the research setting. Self-administered symptom scales have been developed (Hamer et al, 1996) but again these have not found wide usage, in contrast to the emphasis on patient-rated quality of life as an outcome. Clinicians increasingly seek treatment outcomes such as degree of independent living, time to discontinuation of medication, and time to relapse and rehospitalisation rather than changes in symptom rating scale scores (Tiithonen et al, 2006).

**Concurrent validity**

The question remains whether any rating scale (or factorial components of it) demonstrates sufficient concurrent validity to predict these external outcome variables. Operational definitions of remission may achieve this. These consist of multiple item threshold rather than factorial scores, with the addition of a duration condition. In the absence of concurrent validity with other outcome measures, symptom rating scales can only constitute a small part of the appraisal of overall outcome. Symptom rating scales will answer the question ‘Did the antipsychotic drug work on this patient’s symptoms?’ as opposed to ‘What is this patient’s outcome?’ Marshall et al, 2000 emphasise that the use of unpublished rating scales in controlled trials is associated with consistent claims of superiority of new treatments and that familiar, well-validated scales may give a more accurate answer.

**THE BIG THREE**

Three symptom rating scales have dominated the field of schizophrenia research and, in particular, studies of antipsychotic efficacy. With the admonition of Marshall (Marshall et al, 2000) in mind they will be dealt with in some detail here.

**Brief Psychiatric Rating Scale**

The Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962) is a one-page, 16- or 18-item rating scale which was developed more than 40 years ago. It assesses a range of psychotic and affective symptoms rated from both observation of the patient and the patient’s own report. The original purpose of the BPRS was the rapid evaluation of clinical change irrespective of origin (e.g. natural remission or treatment response) in the broad range of psychiatric patients, not just those with schizophrenia. It was not, therefore, specifically designed as an outcome measure; the authors hoped that the scale would develop into a diagnostic instrument, which they considered of greater long-term value than detecting change. Standard definitions of outcome were developed later, e.g. ‘consumer outcome is the effect on a patient’s health status attributable to an intervention by a health professional or health service’ (Andrews et al, 1994). Even so, the authors later stated that the BPRS was designed to fill a special need in clinical psychopharmacology research, at the inception of the Early Clinical Drug Evaluation Units of the National Institute of Mental Health in the USA (Overall & Gorham, 1988).

**Extent of use and adaptation**

The BPRS has perhaps been used more extensively than any other symptom rating scale, in many diagnostic groups and for a wide range of purposes. It is highly sensitive to change, and excellent interrater reliability can be achieved with training and a standard interview procedure (Overall & Rhoades, 1982). As well as the evaluation of efficacy of several classes of psychotropic medication (Hedlund & Vieweg, 1980; Overall & Rhoades 1982; Perry et al, 1997; Hamilton et al, 1998), the BPRS has been used extensively to compare diagnostic concepts internationally and in epidemiological studies (Delmonte et al, 1970; Engelsmann & Formankova, 1967; Engelsmann et al, 1970; Overall & Beller, 1984). It has been translated into many languages and frequently modified for specific purposes, including for use with children (Overall & Pfefferbaum, 1982; Emstley et al, 1997). It has been expanded to 24 items to make it more comprehensive in the area of psychotic and affective symptoms, with items on bizarre behaviour, suicidality, self-neglect, elevated mood, distractability and motor hyperactivity (Ventura et al, 2000). The BPRS has been demonstrated as reliable for use by nursing staff, increasing its utility (McGorry et al, 1988). Most adaptations of the BPRS use one of two scoring versions for each item (either a 0- to 3-point or a 0- to 7-point scale).

**Limitations**

The factor structure of BPRS responses depends upon the characteristics of the patient group under study, and the version being used. The BPRS was, until the advent of the Positive and Negative Syndrome Scale (PANSS; Kay et al, 1987) which itself is partially derived from the BPRS, the most widely used scale in schizophrenia research. This reflected its broad coverage of typical schizophrenia phenomena in the positive, negative and disorganisation categories. However, its coverage of the negative syndrome has been criticised; there are only three negative syndrome items, and it has been suggested that a more extensive scale is necessary for sensitivity to change (Eckert et al, 1996). The authors themselves were dismissive of the use of their scale to determine differences between specific symptoms or...
syndromes during treatment, stating that ‘Although psychiatric symptomatology is multidimensional, the difference between pre-treatment pathology and post-treatment pathology (or lack of it) can be represented by a single dimension spanning the multivariate space’ (Overall & Gorham, 1988). Despite this, with the assistance of 20 psychiatrists, they gave 13 different weights to each item according to diagnosis, in order to increase or reduce the relevance of treatment effects to the total score. For instance, the score on item 8, ‘grandiosity’, would be multiplied by 0.5 in a patient with depression and by 3 in a patient with paranoia. This complex and somewhat arbitrary scoring system appears never to have been taken up.

**Clinical Global Impression**

The CGI is not strictly a symptom rating scale but is included because of its wide use, influence and the recent development of forms specific to the schizophrenia syndromes (CGI-SCH). The original version is a simple instrument which rates the overall severity of any mental disorder (Guy, 1976). This is rated entirely according to clinical judgement in routine professional practice, on a scale for the overall current severity of symptoms from 1 (healthy, not ill) to 7 (among the most severely ill). There is also a 7-point scale for global improvement (usually from baseline to the current condition), rating from 1 (very much improved) to 7 (very much worse). The CGI has been used in several efficacy and effectiveness studies in schizophrenia, is sensitive to change and correlates well with changes assessed with more complex scales (Harro *et al.*, 2003; Leucht & Engel, 2006; Leucht *et al.*, 2006; Rabinowitz *et al.* 2006).

The main criticism levelled at the CGI, that it lacks standard definitions (Beneke & Rasmus, 1992), reflects what many consider its main strength – the use of an adequate level of clinical judgement. Its brevity, utility and appeal to clinical commonsense have ensured its continued use over many more complex rating scales. The CGI has been adapted for the assessment of bipolar affective disorder (CGI–BP) and schizophrenia (Spearing, 1987; Harro *et al.*, 2003). The CGI-SCH has demonstrated good reliability and validity in the evaluation of severity of positive, negative, depressive and cognitive symptoms, and is recommended for both research and clinical practice.

**Positive And Negative Syndrome Scale**

The PANSS (Kay *et al.*, 1987, 1988, 1989) originated from a growing need to reduce the heterogeneity of what was known about schizophrenia. Crow’s (Crow, 1980) positive-negative dichotomy presented a promising theoretical model for explaining and understanding variability in the aetiology of schizophrenia, treatment and prognosis. However, attempts to utilise the model in practice met with inconsistent results (Andreasen, 1982; Andreasen & Olsen, 1982; Pogue-Geile & Harrow, 1984; Lindenmayer *et al.*, 1986), and it was suggested that this might be because of the lack of a comprehensive rating scale for positive and negative symptoms that was feasible, accurate, well validated, reliable, sensitive and standardised. The PANSS, therefore, was not developed to assess outcome per se, or even the results of treatment interventions.

**Nature and scoring**

The PANSS is a 30-item 7-point (1–7) rating scale which amalgamated the 18-item BPRS and 12 items from the Psychopathology Rating Schedule (Singh & Kay, 1975). The items were precisely defined, as were anchor points for the numerical rating of each item. The PANSS was divided into positive, negative and general psychopathology sub-scales (a ‘manic’ sub-scale was later derived; Lindenmayer *et al.*, 2004) and trialled on over 100 well-characterised patients with chronic illness. Sub-scale scores were shown to be normally distributed and independent of each other; they were robust to the effects of mood, chronicity, medication side-effects and cognition. The PANSS was furthermore sensitive and specific regarding pharmacological manipulation of the levels of both positive and negative symptoms in patients with schizophrenia. The validity of its sub-scales was confirmed in an exploration of a classification of patients by predominant symptom class. Sub-scale scores were associated with a number of clinical, treatment and cognitive variables, including premorbid adjustment (Krauss *et al.*, 1998), but not outcome. One of the strengths claimed for the PANSS is consistency in scoring individual patients over time and illness course. A potentially confusing feature of the PANSS, however, is that even those without any mental ill health will score 30. In effect, this means that 30 must be subtracted from the patient’s score in order to gain a meaningful understanding.

**Correlates and factors**

Several studies have sought correlations between PANSS total and sub-scale scores, and other aspects of the illness, to demonstrate concurrent validity. Other aspects have included ventricular enlargement and cortical atrophy (d’Amato *et al.*, 1992), work performance (Bell *et al.*, 1992), neuro-psychological impairment (Bell *et al.*, 1994; Liu *et al.*, 1997; Mass *et al.*, 2000; Bozikas *et al.*, 2004; Good *et al.*, 2004; Ritsner *et al.*, 2006) and violent behaviour (Steinert *et al.*, 2000). Overall these findings appear not to be sufficiently convincing as to be of clinical use, and PANSS scores have generally not been used as proxy variables. For example, when PANSS ‘cognitive’ items were used to predict global cognitive function 66% of the variance was unexplained, suggesting that the PANSS lacked sensitivity and specificity in this regard (Good *et al.*, 2004). This approach appears not to have generated further research hypotheses.

Factorial validity (the nature and purity of the syndromal components of the scale) is essential to the success of investigations utilising sub-scale scores. There are many reports on the factor (syndrome) structure of PANSS items, with much controversy over whether data best fit a three-, four-, five- or even six-factor solution (Peralta & Cuesta, 1994; Lindenmayer *et al.*, 1994; Wolthaus *et al.*, 2000; Fresan *et al.*, 2005; White, 2005; Van den Oord *et al.*, 2006). The simplest factor solutions comprise a syndrome made up of negative symptom items (psychomotor poverty syndrome), a syndrome made up of delusions and hallucinations (reality distortion syndrome) and a syndrome made up of thought disorder and inappropriate affect symptom items (dysorganisation syndrome). Although several five-factor models have been proposed, none has been validated by confirmatory factor analysis (van der Gaag *et al.*, 2006a). This might reflect the ambiguous definitions of some symptom items, such as lack of judgement and insight, which have more than one cause in schizophrenia.

Another complication is that the depression sub-scale (unlike the Calgary Depression Scale; Addington *et al.*, 1992) is unable to distinguish between depression, negative symptoms and extrapyramidal side-effects (Collins *et al.*, 1996). Negative factor scores have been found to correlate
with an independent depression rating instrument (Montgomery Asberg Depression Rating Scale), although depression factor scores did as well (Wolthaus et al., 2000). The loading of single items by multiple causes, which was suggested in another study (Van den Oord et al., 2006) was confirmed in a statistically novel analysis (van der Gaag et al., 2006b).

Only if syndromes possess concurrent validity with other aspects of schizophrenia such as cognitive impairment and poor social function, and furthermore fit explanatory data, can they represent clinical reality. The implication for the rating scale is that items which load on more than one factor must be replaced by two or more items, each of which load on a single factor, which results in lengthier scales. The alternative is losing data through deletion of such items. Poor fit suggests that correlations between syndrome scores and other illness variables under investigation, including outcome, might be unreliable.

MEANING OF SYMPTOM RATING SCALE SCORES

The existence of apparently rival rating scales can be confusing when they purport to measure the same thing. Despite the caveats regarding factorial purity which have been repeatedly addressed in the case of the PANSS, there appears to be much redundancy both within and between rating scales. For example, there are high correlations between positive and negative syndrome scores on the PANSS, and Andreasen’s Schedule for the Assessment of Positive Symptoms (SAPS; Andreasen, 1984a) and Schedule for the Assessment of Negative Symptoms (SANS; Andreasen, 1984a,b; Norman et al., 1996). The negative symptoms of the PANSS and BPRS, and the SANS all measure, mostly, affective flattening rather than the full range of negative symptom phenomena (Welham et al., 1999). The much shorter and quicker CGI scales were just as good as the BPRS in discriminating between the effects of antipsychotic drugs (Leucht & Engel, 2006) despite having been criticised on semantic, logical and statistical grounds (Beneke & Rasmus, 1992). The development of the CGI–SCH scale suggests that investment in less complex rating instruments is gathering pace for rating severity and treatment response in routine clinical practice (Haro et al., 2003).

Even in randomised placebo-controlled trials for licensing purposes, the use of changes in rating scale scores may lack good face validity. Many trials evaluate clinical response as a percentage change in scores over the treatment period. Equating a 20% improvement in symptoms with response follows the study of Kane et al. (1988) which compared clozapine and chlorpromazine in treatment-resistant patients with severe illness. This relatively low percentage reflects the fact that in patients with severe illness even a fairly small attenuation of symptoms might be clinically valuable. The 20% definition of response might not, however, be generalisable to the majority of acute trials with non-resistant patients. Relying on percentage point change to indicate recovery ignores the importance of baseline levels. A 20% reduction of a PANSS score of 100 is double a 20% reduction of a PANSS score of 50, yet both might be recorded as a ‘clinical response’. The patient with a baseline PANSS score of 100 would, although fulfilling criteria for response with a score of 80, remain severely ill, (albeit noticeably less so), whereas the patient with a baseline score of 50 would remain mildly ill with a score of 40 and perhaps not even be noticeably different.

**Concurrent validity**

Leucht et al., 2005a addressed the issue of what rating scale scores mean in clinical terms. They used an equating procedure to anchor BPRS scores to CGI categories (both severity and improvement) across seven drug trials which used both scales in patients with acute schizophrenia. Clinician-rated ‘minimal improvement’ on the CGI equated to a 30% improvement on the BPRS (substantially greater than the generally accepted standard for response). ‘Much improvement’ after 4 weeks of treatment equated to a fall in the BPRS score of almost 58% (Table 1). In addition they found that clinicians used only a small part of the BPRS score range of 18–126: patients with minimum illness on the CGI scored 31, those with moderate illness scored 41 and those with severe illness 53. This is probably because patients are only assessed on a minority of the items and upon most they are scored zero.

Using the same approach with the PANSS (Leucht et al., 2005b) they found that ‘mildly ill’, ‘moderately ill’, ‘markedly ill’ and ‘severely ill’ according to the CGI equated to total PANSS scores of 58, 75, 95 and 116 respectively (Table 2). At 6 weeks, to achieve CGI ratings of ‘minimally improved’ and ‘much improved’ the PANSS decrements were 28% and 53%. The authors suggested that response ought to be defined as a 50% improvement in PANSS score, although in treatment-resistant groups a decrease of 25% might suffice.

A later study (Leucht et al., 2006) compared the PANSS and BPRS with each other and with the CGI and replicated the findings overall, emphasising that smaller absolute score reductions equated to perception of improvement in patients with severe illness compared with those with mild illness (Table 3). For a reduction of 1 point on the CGI Severity of Illness scale there were decreases of 15 and 10 on the PANSS and BPRS respectively.

A similar study (Cramer et al., 2001) found that clinician-rated ‘improved’ and ‘much better’ patients had PANSS scores lowered by 21 and 45% respectively. Quality of life scores were also increased by similar degrees (26 and 50%). This is consistent with the Leucht et al. (2006) study, and perhaps demonstrates some concurrent validity of the PANSS with subjective quality of life as an outcome. A further report indicated that a decrement of 20% on the PANSS equated to a 1-point severity decrease on the CGI-SCH (Rabinowitz et al., 2006).

**Table 1** Clinical implications of BPRS scores

<table>
<thead>
<tr>
<th>Severity of illness</th>
<th>Corresponding BPRS score</th>
<th>CGI Global Improvement</th>
<th>Corresponding BPRS reduction at 1, 2 and 4 weeks, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mildly ill</td>
<td>31</td>
<td>Minimally improved</td>
<td>24, 27 and 30</td>
</tr>
<tr>
<td>Moderately ill</td>
<td>41</td>
<td>Much improved</td>
<td>40, 53 and 58</td>
</tr>
<tr>
<td>Markedly ill</td>
<td>53</td>
<td>Very much improved</td>
<td>71, 79 and 85</td>
</tr>
</tbody>
</table>

BPRS, Brief Psychiatric Rating Scale; CGI, Clinical Global Impression.
REMISSION

These practical difficulties in the use of symptom rating scales to evaluate outcome in treatment trials have contributed to the recent development of the concept of remission in schizophrenia. Response to treatment focuses on short-term improvements and gives little guidance to clinicians regarding long-term management. In general medicine, remission implies a low level of symptoms but with functional recovery. A number of disparate definitions of remission in schizophrenia have been constructed (Leucht & Lasser, 2006). A standard definition, it is argued, is potentially useful: it is realistic and establishes a meaningful treatment goal. Although a useable measure will not include cognition, personal and social function because of difficulties in measurement, there is some evidence that concepts of remission based on symptoms and duration are indeed associated with such consequential aspects of patients’ well-being (Birsoy, et al., 2006).

The Remission in Schizophrenia Working Group was convened in April 2003 to develop a consensus definition of remission in schizophrenia (Andreasen et al., 2005). Taking precedents in physical medicine and affective disorder, remission should be defined as low or mild symptom levels (which by definition do not influence behaviour) and which should last for a minimum, defined duration. Such a standardised definition, unlike several previous published definitions, could be applied across treatment studies and would permit immediate, transparent comparison. This approach does, however, require attention to levels of baseline severity across studies.

The Working Group aimed to map the chosen remission symptoms, which had to be rated mild or less, onto the three best validated syndromes of schizophrenia (reality distortion, disorganisation and negative symptoms) and the five DSM-IV criteria for schizophrenia (delusions, hallucinations, disorganised speech, disorganised or catatonic behaviour, negative symptoms, American Psychiatric Association, 1994). They picked appropriate items from the BPRS, the PANSS the SAPS and the SANS (Table 4).

The BPRS, with limited coverage of negative symptoms, was perhaps less useful in determining remission. The Working Group set 6 months as the minimum duration of symptoms remaining mild for the patient to qualify for remitted status.

Use of remission criteria

Remission is already being used in attempts to test efficacy of drugs in ‘head to head’ comparisons by re-analysing existing data (Sethuraman et al., 2005). A study of stable patients using PANSS-based remission criteria demonstrated that nearly 70% were not in remission; 20% achieved remission when switched to depot treatment and 85% of those already in remission remained so a year later on depot (Lasser et al., 2005). Application of the criteria to data from other published studies produced similar findings (Gharabawi et al., 2005; Kissling et al., 2005). In all studies remission was associated with PANSS total and subtotal scores, CGI–SCH scores, functioning and quality of life. Moreover, an analysis of six clinical trials comparing two definitions, one PANSS based and the other BPRS/CGI based, found that achievement of remission using either definition was associated with better quality of life (Dunayevich et al., 2006). This was particularly so if remission was sustained. Nevertheless, total BPRS change score still contributed the greatest part of the variance in quality of life.

Two reviews of the Working Group remission criteria (Nasrallah, 2006; Van Os, et al., 2006) proposed that the definition was conceptually viable and feasible in both clinical trials and clinical practice. Both reviews considered that the use of remission criteria would raise clinical expectations and drive clinical services to achieve and document better outcomes. In clinical trials, the concept should improve the quality of methodology and data reporting, while extending its relevance to cognition and functional outcomes in patients. The advantages of remission derive from adding duration to absolute symptom score thresholds, and avoiding percentage change scores (a hitherto dubious benchmark).

CONCLUSIONS

Symptom rating scales which have been designed to diagnose patients, subdivide patients, define syndromes, track clinical change or evaluate drug efficacy do not lend themselves easily to the assessment of global outcome in schizophrenia. Simply totalling the number of symptoms without reference to the consequences of what is scored, is an empty exercise. Change must be relative to baseline conditions; there are also issues of redundancy, and a lack of concurrent validity with external outcome measures. The effort expended investigating the psychometric properties of scales such as the PANSS appears to have been matched by only limited advances in their utility beyond tracking change. It has yielded little of relevance to aetiology, treatment or prognosis.
Table 4 Proposed items for remission criteria with cross-scale correspondence and relationship to historical constructs of psychopathology dimensions and DSM-IV criteria for schizophrenia

<table>
<thead>
<tr>
<th>Dimension of psychopathology</th>
<th>DSM-IV criterion</th>
<th>SAPS and SANS items</th>
<th>PANSS items</th>
<th>BPRS items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Criterion</td>
<td>Item no.</td>
<td>Criterion</td>
</tr>
<tr>
<td>Reality distortion</td>
<td>Delusions</td>
<td>Delusions</td>
<td>20</td>
<td>Delusions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SAPS)</td>
<td></td>
<td>Unusual thought content</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hallucinations</td>
<td>7</td>
<td>Hallucinatory behaviour</td>
</tr>
<tr>
<td>Disorganisation</td>
<td>Disorganised speech</td>
<td>Positive formal thought disorder (SAPS)</td>
<td>34</td>
<td>Conceptual disorganisation</td>
</tr>
<tr>
<td></td>
<td>Grossly disorganised or catatonic behaviour</td>
<td>Bizarre behaviour (SAPS)</td>
<td>25</td>
<td>Mannerisms/posturing</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>Negative symptoms</td>
<td>Affective flattening (SANS)</td>
<td>7</td>
<td>Blunted affect</td>
</tr>
<tr>
<td>(psychomotor poverty)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Avolition–apathy (SANS)</td>
<td>17</td>
<td>Social withdrawal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anhedonia – asociality (SANS)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alogia (SANS)</td>
<td>13</td>
<td>Lack of spontaneity</td>
</tr>
</tbody>
</table>

BPRS, Brief Psychiatric Rating Scale; PANSS, Positive and Negative Syndrome Scale; SANS, Scale for Assessment of Negative Symptoms; SAPS, Scale for Assessment of Positive Symptoms.

1. For symptomatic remission, maintenance over a 6-month period of simultaneous ratings of mild or less on all items is required.
2. Use of BPRS criteria may be complemented by use of the SANS criteria for evaluating overall remission.

These limitations have led to interest in another perspective on outcome, remission. This is based on accepted practice in medicine and other psychiatric disorders, such as affective disorders, and goes beyond rating scale scores alone. Its utility, however, remains to be seen. There are already indications that remission may be short lived in many patients (Dunayevich et al., 2006). Until recovery can be defined accurately in schizophrenia (Leucht & Lasser, 2006) symptom control, remission and quantified cognitive, personal and social functioning should be used together as measures of treatment outcome. This accepts that outcome has multiple facets, which vary in importance between patients. Symptom rating scales play an important role in overall appraisal of outcome, but should not dominate the picture, which still requires meaningful appraisals of cognition, personal and social functioning.

REFERENCES


---

ANN M. MORTIMER, BSc, MD, MMedSci, FRCPsych, Department of Psychiatry, University of Hull, Hertford Building, Cottingham Road, Hull HU6 7RX. Email: a.m.mortimer@hull.ac.uk

---


Social outcomes in schizophrenia

STEFAN PRIEBE

Background Outcomes reflecting the social situation are widely considered as important in the treatment of people with schizophrenia.

Aims To review concepts of social outcomes in schizophrenia and the corresponding assessment instruments.

Method Non-systematic literature review and reflection on conceptual and methodological issues.

Results Concepts of social outcomes in schizophrenia lack agreed definitions and theoretical models. A fundamental issue is the distinction between objective and subjective indicators. More research has focused on subjective indicators, which are only weakly correlated with objective life situation and show consistent correlations with mood. Various assessment instruments have been developed pragmatically, particularly to measure quality of life and social functioning, and the literature provides extensive data for comparison.

Conclusions Established instruments exist to measure social outcomes in schizophrenia. Their use requires an awareness of the specific strengths and limitations.

Declaration of interest None.

Since the beginning of systematic outcome assessment in schizophrenia in the 1960s, there has been a wide consensus among researchers and clinicians that capturing psychopathological symptoms alone is not sufficient to reflect relevant outcomes. Particularly for evaluating long-term outcomes, information on the social situation of patients is regarded as essential. Social outcomes assess how patients live, function in society and perform their various roles.

Social outcomes are commonly used throughout healthcare. Yet, there are some specific reasons for their popularity in the treatment of schizophrenia:

(a) The disorder is often persistent and affects patients lifelong. Symptoms and the associated distress may fluctuate, and establishing symptoms at any point in time might therefore yield a less relevant picture than the more stable social situation.

(b) Longitudinal research has shown that antipsychotic medication can reduce productive symptoms and prevent relapses with subsequent re-hospitalisation. Yet, this effect was not necessarily linked with an improved social situation. Symptom improvement and prevention of relapses alone do not make patients necessarily more likely to complete education, find employment and have social relationships. These outcomes need therefore to be assessed separately from symptoms.

(c) As a result of mental health reforms in most high-income countries, the focus of care has shifted from the asylum to the community. Former long-term hospitalised patients were discharged, and there was an interest in how they fared in the community without the institutional protection of the asylum.

(d) Mental health reforms have been associated with the formation of patient organisations. Such organisations have acquired an important voice in debates on mental health policies and commonly demand that social outcomes are given more prominence in both research and practice.

Other stakeholder groups often share the perspective of patient organisations. In public and professional debates on mental healthcare, it is often felt that what really matters is how patients live (e.g. whether they do or do not have a job and friends) rather than symptoms of illness.

IMPACT OF TREATMENT ON SOCIAL OUTCOMES

How can treatment of schizophrenia affect social outcomes? There are at least three possible mechanisms:

(a) Treatment can improve psychopathological symptoms. A lower symptom level can enable people with schizophrenia to function and perform better in their social context and subsequently achieve more favourable social outcomes.

(b) Treatment may have an impact not only on conventional psychopathological symptoms, but also on other cognitive and social deficits that are illness related, but are usually not captured in psychopathological assessments (e.g. the concept of social cognition, which has received wide attention in the past 5 years). If treatment diminishes deficits in social cognition, patients might be more likely to establish and maintain useful relationships and improve social outcomes.

(c) Some care interventions focus directly on social outcomes. For instance, vocational rehabilitation programmes may improve the work situation, and the effect is not mediated by a reduction of symptom levels or other illness-related deficits.

Although the latter mechanism mainly applies to a range of social management interventions, the other mechanisms can operate with all forms of psychological, pharmacological and socio-therapeutic treatments. In practice and research, the mechanisms can be intertwined in a complex way and are difficult to disentangle. Yet, it may be concluded that social outcomes can be used to assess the effects of all forms of treatment in schizophrenia.

Because of the indirect nature of the potential treatment effect on social outcomes, they have been termed ‘distal’, as opposed to the more ‘proximal’ outcome criterion
of psychopathological symptoms (Watts & Priebe, 2002). The effect on social outcomes is less immediate than on symptoms, and achieving improvements in a person’s social situation usually takes time.

**SOCIETAL CONTEXT AND CEILING EFFECTS**

For the analysis of treatment effects on social outcomes, two fundamental issues should be considered. One is the dependency of social outcomes on the societal context. For example, the likelihood of a person with schizophrenia obtaining competitive employment as a result of treatment will heavily depend on societal factors such as the general unemployment rate and legislation for the employment of people with disabilities. Thus, social outcomes will rarely be a function of treatment alone. Another issue when using social outcomes for evaluating treatment is their distribution in the treated sample at baseline. Psychopathological symptoms define the illness and will always be at a considerable level at the beginning of treatment, which leaves room for improvement. To some extent, this also applied to social outcomes in many studies when people with schizophrenia were discharged after long-term hospitalisation or began treatment in very unfavourable circumstances. However, there can be exceptions. It is possible that people have symptoms of schizophrenia but at the same time hold a respectable and satisfactory social position and perform well in different societal roles. In such a case, no treatment can improve the social situation. At best it can help to maintain the current level. Thus, unlike psychopathological symptoms, treatment cannot always aim to improve social outcomes, and whether maintaining the given social situation can be rated a success is a difficult question and depends on the quality of the social situation before treatment.

**CONCEPTS OF SOCIAL OUTCOMES**

Different concepts have been used to reflect and summarise social outcomes in people with schizophrenia. These include standard of living, quality of life, social integration, social adaptation, social functioning, social integration, needs for care and, more recently, social inclusion. None of these concepts was introduced into psychiatry on the basis of a theoretical model. If a theoretical literature existed in psychology and sociology – e.g. for the quality of life concept – it was rarely considered when new concepts were suggested and new assessment tools were designed in psychiatry.

The reason for introducing a new concept was commonly the intuitive appeal of the term, which then led to efforts to find definitions and, subsequently, develop corresponding assessment tools. There is no universally accepted definition for any of the concepts, and each can be used and has been used in various ways, depending on the perspective and interest of whoever uses them. Since the 1980s researchers have published definitions and taken a pragmatic and often ad hoc approach to developing operationalised methods for the assessment. The operationalisation usually required some focus and narrowing down of the various potential meanings of the concepts. As a result, there is a tendency that all assessment instruments for social concepts lead to a disappointment in at least some stakeholder groups because they do not exactly reflect the specific or vague understanding of the concept in the given group. To a different degree, this has happened whenever new concepts of social outcomes have superseded previous concepts. Books on quality of life and social functioning, the two dominating concepts, were published in the 1990s (Tyrer & Casey 1993; Katschnig et al, 1997; Priebe et al, 1999b) with limited conceptual and methodological progress since.

**OBJECTIVE AND SUBJECTIVE INDICATORS**

Whatever concept is used in the assessment of social outcomes, there are objective and subjective indicators. Objective indicators are facts about the social situation, which – at least in theory – can be objectively and unequivocally assessed. These include whether a patient does or does not have employment, a partner, independent accommodation and social contacts. Such ‘hard’ outcomes are transparent, straightforward to interpret and of obvious relevance. Out of all outcome criteria in schizophrenia, they arguably have the highest appeal to the public and user organisations. If a treatment has a demonstrable positive effect on the employment of patients, to most stakeholders this will be more persuasive of its value than an impact on scales assessing symptoms or other psychological constructs.

Objective indicators are important, widely accepted and relatively easy to establish. Why is it then that they are not more widely used and reported in studies on the outcome of treatment in schizophrenia? There are several reasons:

(a) Objective aspects of the social situation are hard to influence and are very ‘distal’ outcome criteria. For example, pharmaceutical companies might argue that influencing the objective social situation is too ambitious an aim for treatment with antipsychotic medication, that demonstrating an impact on the objective social situation would take much longer than the usual length of clinical trials, and that such a criterion would be inappropriate because pharmacological treatments were developed to reduce symptoms, not as ‘employment-finding’ drugs.

(b) Objective indicators tend to be difficult to change. Even over longer periods people with schizophrenia will not easily move into competitive employment, find a partner and achieve independent living. In a larger sample some might improve on any one of these criteria, but seldom on all. For meaningful statistical testing of changes over time, the different aspects would have to be combined to have a sufficient frequency of changes and to avoid multiple testing. Interestingly, such a combined measure of objective indicators does not exist.

(c) Any outcome criterion may have problems with floor and ceiling effects but this applies particularly to hard social outcomes. People who already are in independent living and competitive employment cannot improve any more, so that these criteria become meaningless as outcomes.

(d) Although objective indicators capture hard facts and are usually straightforward to analyse, their interpretation requires values, and these values can vary depending on cultural background, social context and individual perspectives. For example, for most people being in employment is clearly desirable, but how does one assess social outcome in a person who does not want to work and can afford to live on other income? The dependency of the assessment on values is even more obvious with respect to partnership and social contacts. People
might choose to live alone rather than being forced into this as a result of illness-related impairment. One solution to this dilemma is to ask patients about their expectations and aspirations, and relate their social situation to their wishes. Following this approach, social isolation would be a negative outcome only if the person would prefer to have more contacts. This, however, goes beyond objective indicators and introduces a subjective dimension.

Subjective indicators comprise patient ratings of feelings, thoughts and views on their social situation. An appropriate description of the full range of social indicators used in different concepts is beyond the scope of this review, but it will focus on quality of life, which is the most frequently used concept in social outcomes in the psychiatric literature.

**QUALITY OF LIFE**

Since the 1980s, quality of life has been increasingly used as an outcome criterion in psychiatric research. Commonly, objective and subjective indicators are considered. Lehman et al. (1982) introduced a measurement approach, which assesses personal characteristics, objective indicators in different domains of life and subjective quality of life in the same life domains. Subjective quality of life represents the person's appraisal of their objective life conditions, mostly captured by rating scales of satisfaction with life domains and life as a whole. The life domains covered usually include work, accommodation, family, social relations, leisure, safety, finances, and physical and mental health. The mean score of the satisfaction ratings – or similar subjective ratings – is taken as the level of subjective quality of life (Priebe et al., 1999a).

Patients' appraisal of their life is influenced by three major processes: a comparison with original expectations and aspirations; a comparison with the life situation and achievements of others; and an adaptation over time. The latter two may be particularly relevant for people with chronic schizophrenia, whose peer group is often people with similar impairments, and who may adapt to circumstances that they might have found unsatisfactory many years earlier. As a result, people with persistent disorders who often live in conditions that seem adversarial and unpleasant to clinicians and observers, nevertheless express relative satisfaction with their life (Arns & Linney, 1993; Awad et al., 1997; Katschnig et al., 1997; Priebe et al., 1999b).

Correlations between objective and subjective indicators are reported to be weak to moderate (ranging from 0.04 to 0.57; Priebe & Fakhoury, 2007). The low association between objective life situation and patients' subjective appraisal has often been counterintuitive to clinicians and other observers, who subsequently questioned the validity of patient ratings. Yet, if patients are asked to give a subjective appraisal of their situation and express a high satisfaction with how they live, there is hardly any external criterion based on which such an appraisal may be disqualified. Thus, patients' views and satisfaction ratings may look surprising to the independent observer, but need to be respected as subjective indicators.

**Assessment instruments**

A range of scales, checklists and structured and semi-structured interviews have been developed to assess quality of life in people with schizophrenia. The results of scales assessing symptom levels, particularly of depression, have been repeatedly reported as quality of life scores, although the scales have been developed neither to assess quality of life nor to capture objective and subjective indicators. These proxy measures will not be considered here. Table 1 shows a number of established scales that have been specifically developed to assess quality of life and have been used in people with schizophrenia. The listed assessment instruments for quality of life – and later social functioning – were identified through a non-systematic and non-exhaustive literature search and were selected on the basis of their use in research.

To assess quality of life in people with schizophrenia, generic, health-related and disease-specific instruments can be used. Generic scales can be applied to the general population and any group of people with health problems, including schizophrenia. Scales often include questions on physical and mental health, but these are not specific to any illness or treatment. Results can be compared across groups with different characteristics and disorders, irrespective of the type of intervention received. Examples are the Quality of Life Interview (QLI; Lehman, 1983), the Lancashire Quality of Life Profile (LQOLP; Oliver, 1991) and the Manchester Short Assessment of Quality of Life (MANSA; Priebe et al., 1999a).

Health-related quality of life measures are targeted to assess the quality of life of samples with health problems irrespective of the type of illness and interventions. Examples are the Medical Outcome Study Questionnaire (MOS), which was modified and shortened to the 36-item Short-Form General Health Survey (SF–36; Ware & Sherbourne, 1992), and the EuroQOL–5D (EQ–5D; EuroQol Group, 1990). There are also disease-specific measures, and several of these have been designed to assess the quality of life of people with schizophrenia. A widely used disease-specific instrument is the Quality of Life Scale (QLS; Heinrichs et al., 1984), which is a clinician rating scale with acceptable psychometric properties. It was developed to assess symptom levels and functional status of people with schizophrenia in longitudinal studies and trials. Other, less widely used examples of disease-specific scales are the Subjective Well-being under Neuroleptics Scale (SWN; Naber, 1995) and the Schizophrenia Quality of Life Scale (SQLS; Wilkinson et al., 2000). These scales tend to capture symptoms, in particular mood symptoms, and side-effects of antipsychotic medication. Although they may be important in influencing quality of life, the labelling of these factors as quality of life is questionable and can blur the concept. It might be preferable to measure symptoms as symptoms and side-effects as side-effects, instead of declaring them to be a direct indicator of quality of life.

There are differences in the use of the scales worldwide. In the USA, the QLS, QLI and the Oregon Quality of Life Scale (OQLS; Bigelow et al., 1991) have been more widely used, whereas in Europe the LQOLP, the MANSA and the EQ–5D are more popular. The previous use of an instrument and the availability of data for comparison are powerful determinants of the choice of instrument. Other determinants are the time to complete the instrument, the requirements for training, the properties of the instrument, its overall approach and exact content, and the purpose of the data collection.

**Properties of instruments**

With respect to instrument properties, the literature usually reports psychometric
characteristics such as validity, reliability and objectivity. Yet, these terms are based on psychological test theory and the assumption that there is a well-defined construct that needs to be measured. In the assessment of social outcomes, one might argue that there is no well-defined concept and psychological test theory does not apply. Are social outcomes tested or are they simply assessed and documented? In the latter case, assessment tools are methods to document objective indicators and patient statements. The results on each question can – unlike in psychological test theory – be directly interpreted. Items can be summarised in scores, but the score does not necessarily reflect an underlying construct. To be administered usefully in longitudinal assessments, scales still need to have certain qualities, such as providing stable results over time in the absence of changes in the person’s social situation. Yet, this would not be a conventional test–retest reliability because there is no construct to be ‘tested’. This is a theoretical debate which, however, is important for interpreting results of social outcome measures, and should be addressed in the future development of new instruments.

SOCIAL FUNCTIONING

After quality of life, social functioning of people with schizophrenia has received the most extensive attention in the psychiatric literature. Instruments assessing social functioning capture the capacity of a person to function in different societal roles and their actual social performance. Table 2 shows instruments to assess social functioning in people with schizophrenia.

As in quality of life assessment, the selection of an instrument depends on various factors, and an ideal scale for all purposes does not exist.

EMPIRICAL FINDINGS

The literature on social outcomes in general in people with schizophrenia and on their quality of life specifically is vast. Some results cast light on the strengths and weaknesses of social outcomes, in particular subjective indicators.

Subjective quality of life is less favourable in people with schizophrenia who are younger, male, live alone or are homeless, have a high level of education and are not employed (Lehman et al., 1995; Priebe et al., 1998; Priebe & Fakhoury, 2007). Yet, these characteristics explain only a small amount of the variance of subjective quality of life scores in clinical samples. The most consistent and relevant factor influencing subjective quality of life in people with schizophrenia is the level of psychopathological symptoms, in particular mood. The more depressed a person is the lower the subjective quality of life. This applies to both cross-sectional and longitudinal associations (Kaiser et al., 1997; Priebe et al., 2000). The causality of the association, however, is not straightforward. Depression may lead to a negative appraisal of life, and, vice versa, a negative experience of the life situation may lead to depression. Also, both depression and negative appraisal may be symptoms of the same underlying cognitive and affective processes. In any case, an assessment of subjective indicators of social outcomes needs to control for mood as a potential confounding factor.

Social outcomes have been used widely to evaluate the effects of different treatment methods in schizophrenia. Although programmes aimed at improving the social situation directly, such as vocational rehabilitation and discharge from long-term hospitalisation (Priebe et al., 2002), can have a substantial effect, such an impact has only rarely been demonstrated for established pharmacological and psychotherapeutic interventions (Corrigan et al., 2003; Wiersma et al., 2004).

CONCLUSIONS

Social outcomes have a high intuitive appeal and are called for by different stakeholder groups, including the public and user organisations. However, established scales to assess social outcomes lack a
The findings of interest. Their strengths and weaknesses have been well documented. On balance, they should be used to assess outcome and capture the central view of the patients concerned. To use them appropriately, there are at least three requirements: (a) whoever uses such concepts should be aware of the limitations and have a good understanding of what the selected instrument actually assesses, independent of the title of the scale; the contents of scales need to be considered along with practical aspects, when the best instrument for the given purpose is selected; (b) it is difficult to justify the use of more than one instrument to assess subjective indicators of social outcomes in the same study; what the scales assess is conceptually not distinct, and scores of different instruments overlap (Fakhoury et al., 1997); (c) symptom levels, and in particular mood, need to be assessed and controlled for in any analysis of patient ratings of social outcomes.

Rather than taking new appealing terms and pragmatically developing scales to assess them, future work on improving assessment tools should be based on defined theoretical models and take the existing empirical findings into account.

Within psychiatry, schizophrenia research has led in the development of methods to assess social outcomes. It is widely seen as mandatory to assess social outcomes in epidemiological studies and clinical trials. The literature provides sufficient evidence for the use of assessment instruments and appropriate interpretation of the results. Yet, despite several decades of research, more needs to be done to specify the concepts and develop better assessment instruments. This requires approaches that are qualitatively new, and not just more of the same.

REFERENCES


s20


Patient-reported outcomes in schizophrenia

ROSEMARIE McCabe, MARYA SAIDI and STEFAN PRIEBE

Background  Patient-reported outcomes are increasingly used to evaluate the care of people with schizophrenia.

Aims  To review established and emerging patient-reported outcomes in schizophrenia research, assessment tools and key findings.

Method  A non-systematic review addressing relevant constructs, the associated scales and key empirical findings.

Results  Patient-reported outcomes in schizophrenia relate either to evaluation of illness and benefit from treatment or to resilience of the self. Of the former, needs for care, treatment satisfaction and the therapeutic relationship are most common. Less common are symptoms, insight, attitude towards medication, and clinical communication. Increasing expectations of treatment have led to new measures assessing resilience of the self, including empowerment, self-esteem, sense of coherence and recovery. Scores of different patient-related outcomes overlap and are influenced by a general tendency, largely influenced by mood, for more or less positive appraisal.

Conclusions  The conceptual and empirical basis for different patient-reported outcomes varies, with most data available for treatment satisfaction. More than one such outcome should be used only if there is a specific hypothesis. For new patient-reported outcomes, relative independence from existing constructs should be demonstrated.

Declaraton of Interest  None.

Patient-reported outcomes are widely used in mental health research to assess treatment benefits for patients. They are defined by the United States Food and Drug Administration (2006) as ‘any report coming directly from patients (i.e. study subjects) about a health condition and its treatment’. Thus, a patient-reported outcome is any outcome based on a patient’s perception of a disease and its treatment(s) scored by the patient, without any interpretation by a clinician or researcher. A patient-reported outcome can be assessed through single-item or multi-item measures and provides a means of assessing treatment benefit by capturing concepts related to how a person feels or functions with respect to their health status. We use the term patient-reported outcome synonymously with subjective evaluation criterion, that is a criterion for evaluating care based on and directly reflecting the patient’s views, feelings and judgements.

Since the 1970s patient-reported outcomes have become increasingly important for the evaluation of treatment of people with schizophrenia. First, some treatment effects are known only by the patient and hence cannot be measured by observers. This is especially the case in psychiatry as most symptoms cannot be readily observed (e.g. paranoid thoughts) and are not accompanied by physical signs. Second, patients provide a unique perspective on treatment effectiveness. This is particularly important when improvements in clinical measures may not correspond to improvements in how the patient feels or feels. For example, some patients report that the side-effects of antipsychotics are so bad that they would rather hear voices. Third, self-rated instruments might be more reliable than observer-rated measures because they eliminate interrater variability. Fourth, and probably most importantly, the use of patient-reported outcomes reflects the role of the patient as the ‘consumer’ of care. Although traditionally people with schizophrenia may have been treated as passive recipients of treatment, over the past five decades they have increasingly been seen as active partners in care whose views and opinions matter (Priebe et al, 1998).

From the patient perspective, capturing psychopathological symptoms only is not sufficient to reflect relevant outcomes. Improvements in outcomes related to functioning and well-being are also important dimensions of successful treatment (Fleischhacker et al, 2005). Fischer et al (2002) conducted focus groups with patients and identified six goals of treatment: increasing energy and interest; improving social relations; reducing disturbing or unusual experiences (hallucinations and delusions); reducing confusion and difficulty concentrating; reducing medication side-effects; and increasing productive activities such as having a job. In a follow-on study of the outcome priorities of people with schizophrenia, Rosenheck et al (2005) found the strongest preference was for reducing confusion and increasing energy and the least for improving social life and reducing side-effects. However, preferences depended on patients’ well-being and clinical status. Weller patients were more interested in recovery-oriented goals such as social relations, employment and personal energy whereas those who were less well were more concerned with symptoms, confusion or side-effects. Hence, patients’ preferences and priorities for improvement are not uniform and depend on their current clinical status. Regardless of specific preferences, it is clear that patient-reported outcomes are increasingly accepted and used in both research and routine clinical care.

This review presents an overview of patient-reported outcomes in the context of schizophrenia. It will present the underlying constructs, the corresponding scales, touching on their psychometric properties, and key empirical findings relating to these constructs.

This review is necessarily selective owing to the burgeoning pool of relevant constructs and associated measures in mental health research. It will focus on two broad groups of outcomes: those relating to evaluation of illness and treatment and those relating to the patient’s psychological well-being. Although there is no existing conceptual framework within which to classify patient-related outcomes, the former are constructs which could be described as emanating...
from a more clinical perspective, ranging from how the patient views their illness to their perspective on the quality of care provided. The latter are more psychotherapeutic in orientation, being concerned with the well-being of the individual. Patient-reported outcomes commonly used in evaluating treatment (see Table 1) are met and unmet needs for care, treatment satisfaction, and the therapeutic relationship. Less common outcomes rated by the patient, for differing reasons, are symptoms, insight, knowledge about illness/medication, medication side-effects, and the quality of clinical communication. Nasrallah et al. (2005) have written about raising expectations about the prospects of functional recovery among people with schizophrenia. This is reflected in new constructs and measures in mental health research that evaluate psychological well-being of the self (see Table 2). They are shifting the emphasis from symptom management to maximising the person’s quality of life, empowerment, self-esteem, sense of coherence and recovery.

OUTCOMES RELATING TO ILLNESS AND TREATMENT

Needs for care

Constructing a care plan based on assessment of an individual’s needs is fundamental to community care (Brewin et al., 1987). Studies show that needs for care are often assessed quite differently by patients and mental health professionals (Hansson et al., 2001), with disagreements about the number of unmet needs and the area of need (Slade et al., 1998). This is important because unmet needs are strongly associated with quality of life (Slade et al., 1999). The most commonly used measures are the Camberwell Assessment of Need (CAN; Phelan et al., 1995) and the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS; Slade et al., 1999). The CAN assesses perceived need in 22 different areas of life (e.g. accommodation, self-care, daytime activities, intimate relationships) and whether patients are currently receiving any effective help with these difficulties. It can be used to assess the perceptions of the patient, their carer and a member of staff working with them. The CANAS is a shortened version of the CAN covering the same areas. Buhler et al. (2001) reported that people with schizophrenia can validly estimate their needs, and better executive functioning may be associated with the ability to get one’s needs met, increased awareness of needs, better ability to communicate needs, or more needs in certain areas. In six European countries, Kovess-Masféty et al. (2006) found that on average one in four patients had needs (approximately 6 per patient) that were not adequately met by their mental health service.

Treatment satisfaction

Treatment satisfaction in schizophrenia has been used broadly to assess satisfaction with treatment as a whole and also more narrowly to assess satisfaction with antipsychotic medication. It is central to treatment adherence (Chue, 2006). Measures of treatment satisfaction include the Verona Service Satisfaction Scale (VSSS; Ruggeri & Dell’Agnola, 1993), the Client Assessment of Treatment (Priebe et al., 1995) and the Client Satisfaction Questionnaire (CSQ; Artiksson & Zwick, 1982). The 82-item VSSS addresses seven dimensions: overall satisfaction, professionals’ skills and behaviour, information, access, efficacy, types of intervention and relative’s involvement. The dimension ‘professionals’ skills and behaviour’ appears to be the most significant contributor to satisfaction (Henderson et al., 2003). De Wilde & Hendriks (2005) hypothesised that before-treatment self-reported and observer-rated problem severity and treatment need might have significant effects on satisfaction. No relationships were found, however, supporting the hypothesis that the CSQ is primarily influenced by treatment variables rather than patient characteristics and is thus a good indicator of quality of treatment.

In most samples and treatment settings, mean scores of satisfaction with treatment are positive, and differences between the satisfaction with distinct forms of treatment are rather small or self-evident (e.g. patients are less satisfied with restraint and seclusion on wards). This has led to criticism of the construct. Yet, it is the most commonly used patient-reported outcome, and randomised controlled trials on new treatment methods have repeatedly found statistically significant gains in satisfaction with treatment among people with schizophrenia.

Subjective response to and satisfaction with antipsychotics are important to assess because psychiatrists tend to underestimate the level of distress resulting from side-effects (Day et al., 1998). The Drug Attitude Inventory (DAI; Awad, 1993) is a 30-item scale measuring subjective responses to medication (including acceptability and tolerability) which aims to understand the factors influencing adherence. A brief 10-item version (DAI–10) has also been found of use (Hogan et al., 1983). Patients expressing dissatisfaction with their medication tend to have legitimate complaints such as dysphoric reactions and side-effects that have been ignored by their physicians (Hamann et al., 2005). Attitudes toward medication are predicted by insight, the therapeutic relationship with the prescriber and experience of admission (Day et al., 2005), and are associated with adherence to medication and treatment outcome (Awad, 1993).

Therapeutic relationship

Patients view the therapeutic relationship as the most important element of good psychiatric care (Johansson & Eklund, 2003). Their views of the therapeutic relationship are generally quite positive (Priebe & Gryuers, 1993). Empirical evidence shows that a lower degree of satisfaction is normally associated with a higher degree of observer-rated psychopathology (Neale & Rosenheck, 1995; McCabe & Priebe, 2003). Older patients (Drane & Solomon, 1996) and those with more service contacts (Klinkenberg et al., 1999) tend to rate the relationship more positively. Factor analyses of therapeutic relationship scales tend to identify a large global factor accounting for most of the variance (McCabe & Priebe, 2004). Negative ratings of the helping alliance – in one study assessed by a single simple question (‘Do you feel better after talking to your keyworker? Yes or no?’) – have predicted subsequent hospitalisation (Priebe & Gryuers, 1995). These findings suggest that the patient–professional relationship might be a relevant therapeutic factor not only in psychotherapy but also in complex psychiatric treatment settings common in the care of people with schizophrenia (Priebe & Gryuers, 1993).

Most therapeutic relationship scales used in psychiatric research have been developed in psychotherapy (McCabe & Priebe, 2004). However, the Scale to Assess the Therapeutic Relationship (STAR; McGuire-Snieckus et al., 2007) was specifically developed in the context of psychiatric treatment of people with severe mental illness using psychometric test construction principles. It has 12 items making up three
sub-scales: positive collaboration, positive clinician input and non-supportive clinician input. Female patients rated positive clinician input more highly. Patients with higher symptom levels had lower STAR scores and rated clinicians as having more ‘non-supportive clinician input’.

The 30-item Therapist–Patient Relationship Scale with Schizophrenic Patients (Stark et al, 1992) was developed for people with schizophrenia but is limited to capturing aspects of expressed emotion. In the patient version, the patient rates how the therapist relates to them along with the therapist’s therapeutic skills. The Working Alliance Inventory (WAI; Horvath & Greenberg, 1989) was developed in psychotherapy to yield three alliance components (goal, task and bond), and has been used fairly widely in psychiatric research (Bale et al, 2006).

Finally, the Helping Alliance Scale (HAS; also referred to as the Helping Alliance Questionnaire; Priebe & Gruyters, 2001) is a short six-item scale covering the basic elements of therapeutic relationship scales (i.e. feeling understood, respected and receiving the right care). Recently, it has been the most frequently used scale to measure the quality of the therapeutic relationship in research evaluating care of people with severe mental illness. Using the WAI and HAS, Bale et al (2006) found that patient and keyworker scores were only weakly related for assertive community treatment.

<table>
<thead>
<tr>
<th>Table 1 Patient-reported outcomes: assessing illness and treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference(s)</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Camberwell Assessment of Need Short Appraisal Schedule (CANSAS)</td>
</tr>
<tr>
<td>Verona Service Satisfaction Scale (VSSS)</td>
</tr>
<tr>
<td>Client Satisfaction Questionnaire (CSQ)</td>
</tr>
<tr>
<td>Drug Attitudes Inventory (DAI)</td>
</tr>
<tr>
<td>Scale To Assess the Therapeutic Relationship (STAR)</td>
</tr>
<tr>
<td>Therapist–Patient Relationship Scale with Schizophrenic Patients (TPRS)</td>
</tr>
<tr>
<td>Working Alliance Inventory (Client Version)</td>
</tr>
<tr>
<td>Helping Alliance Scale (HAS)</td>
</tr>
<tr>
<td>Approaches to Schizophrenia Communication Scale (ASC)</td>
</tr>
<tr>
<td>Brief Symptom Inventory (BSI)</td>
</tr>
<tr>
<td>Symptom Checklist–90–R (SCL–90–R)</td>
</tr>
<tr>
<td>Insight Scale</td>
</tr>
<tr>
<td>Knowledge About Schizophrenia Questionnaire (KASQ)</td>
</tr>
</tbody>
</table>
Clinical communication

Professional–patient communication is gaining increased attention in medicine generally, but has been relatively neglected in psychiatric research (Hassan et al., 2007). Patient involvement in decision-making is advocated, but concerns about the feasibility of shared decision-making in the treatment of schizophrenia have been raised in relation to the patient’s ability to rationally evaluate treatment options, their severe mistrust of people when they are paranoid and their attention problems (Hamann et al., 2006). Nevertheless, involving patients in their treatment, even when they are acutely ill, is feasible and has been linked with a more positive attitude to medication (Day et al., 2005) and with increased uptake of psychoeducation (Hamann et al., 2006).

Two patient-rated measures of communication, one focusing on needs and the other on side-effects of medication, appear to be useful to highlight aspects of treatment that the clinician might not be aware of or the patient is too embarrassed to raise of their own accord. The Two-Way Communication Checklist (2-COM; Van Os et al., 2002) is a 19-item self-report schedule which highlights areas of need that the patient wishes to discuss with their clinician and has been found most useful by patients with the highest care needs. Perhaps because there was very low concordance between patients and professional carers on individual needs in a naturalistic study, the 2-COM improved doctor–patient communication and led to changes in patient management in a randomised controlled trial (Van Os et al., 2004).

The Approaches to Schizophrenia Communication Scale (ASC; Dott et al., 2001) assesses subjective response to medication with the aim of improving doctor–patient communication and increasing treatment adherence. Respondents identify which of 18 common side-effects of antipsychotics (e.g. difficulty sleeping, impairments in concentration, sexual dysfunction) have troubled them recently and which areas they wish to discuss further with their clinician. Dott et al. (2001) found that 86% of patients found the ASC useful in communicating their problems to members of the healthcare team. Weiden & Miller (2001) found that patients had concerns about sexual functioning without ever having complained to the clinician. When asked why, patients said they had been too embarrassed; it was easier to report this problem on a form, even though they knew that the same clinician would be reading it.

Symptoms

Self-rated symptom measures do not have a strong tradition in the treatment of schizophrenia. As people with schizophrenia have been viewed as having poor insight, the validity of their assessment of their own symptoms has been questioned. This contrasts with a stronger reliance on self-rated symptoms in depression and anxiety. None the less, some studies have used measures of self-rated symptoms. The Brief Symptom Inventory (BSI; Derogatis, 1992) was developed from the Symptom Checklist–90–R; (SCL–90–R; Derogatis, 1983) and is a 53-item scale assessing somatisation, obsessive–compulsive behaviour, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation and psychoticism. Weak-to-moderate correlations have been reported between self- and observer-rated symptoms (Deluty et al., 1986; Fava et al., 1986). Priebe et al. (1998), and Fakhoury et al. (2002) found that self-rated symptoms, self-rated needs and subjective quality of life were significantly correlated in people with schizophrenia.

Insight

The term ‘insight’ is widely used by clinicians to describe a patient’s understanding of their illness. Many definitions and assessments also incorporate attitude towards treatment, in particular, willingness to adhere to prescribed treatment. As such, it is primarily a clinician’s concept and, consequently, most insight scales are rated by a clinician or independent observer. However, there are a few self-rated insight scales. The self-report Insight Scale developed by Birchwood et al. (1994) assesses the three factors proposed by David (1990), namely awareness of illness, need for treatment and attribution of symptoms. Markova & Berrios (1992) developed an insight scale that can be either observer- or self-rated. They broadened the definition of insight to include not only the person’s knowledge about how the disorder affects them but also how it affects their interaction with the world. Less insight is associated with more severe positive symptoms. Patients with more insight have more positive attitudes to medication (Freidenreich et al. 2004). However, greater insight is also related to increased depression and poorer subjective quality of life (Karow & Pajonk, 2006).

The Knowledge About Schizophrenia Questionnaire (Ascher-Svanum, 1999) is related to insight and is used in assessing the outcome of psychoeducation. It is a 25-item multiple choice test to assess patients’ knowledge about their illness and its management. It measures patients’ knowledge about the diagnosis of schizophrenia and its prevalence, aetiology, course and prognosis, knowledge about medication and its side-effects, psychological treatments, stress factors and legal issues. It meets the need to measure the impact of patient education, now prevalent partly because of the growing emphasis on psychosocial rehabilitation and patient empowerment, and helps to demonstrate a meeting of standards of care for patient education in the USA (Ascher-Svanum, 1999).

PSYCHOLOGICAL WELL-BEING AND RESILIENCE OF SELF

Patient-reported outcomes focusing on the person’s psychological well-being reflect a shift from the evaluation of treatment in terms of symptom management to resilience in the face of illness and functional recovery. Constructs that have emerged in the context of schizophrenia research in the past two decades include empowerment, self-esteem, sense of coherence and recovery (Table 2).

Empowerment

Empowerment of patients has its origins in the USA and emphasises patients’ rights to self-determination and their economic situation as consumers of services (Laugharne & Priebe, 2006). It has been suggested that there are two relevant factors: empowerment of the self (higher self-esteem and efficacy) and empowerment within the community (giving the patient greater confidence within the community) (Corrigan & Garman, 1997). The most widely used scale, the Empowerment Scale (Rogers et al., 1997) has five distinct sub-scales: self-esteem, power, community activism, optimism and righteous anger. Age, gender, ethnicity, education, employment and number of hospital admissions were not related to empowerment. Community activism was related to greater empowerment, and use of
services to less empowerment (Rogers et al., 1997). Wowra & McCarter (1999) found that empowerment was influenced by employment status and education level but not race, gender or age.

**Self-esteem**
The disparaging nature of many peoples’ voices and the content of delusional beliefs may affect self-esteem. The Rosenberg Self-Esteem Scale (Rosenberg, 1965), developed in the field of psychology, is comprised of ten items about self-worth. Sorgaard et al. (2002) found that self-esteem was related to mental health and, to a lesser extent, social network. Anxiety/depression and affect balance were the strongest predictors of self-esteem, and having at least one close friend was associated with positive self-esteem. Demographic characteristics played a negligible role, with only gender (female) associated with positive self-esteem. Predictably, negative evaluation of the self was found to be strongly associated with positive symptoms and also with a more critical attitude from family members (Barrowclough et al., 2003). In terms of mental health interventions, Gumley et al. (2006) found that cognitive–behavioural therapy had a positive effect on self-esteem. In addition to viewing poor self-esteem as a consequence of positive symptoms, recent work also suggests a causal role for self-esteem in the development of low mood. Smith et al. (2006) found that people who were more depressed and had lower self-esteem had more severe auditory hallucinations and persecutory delusions with more negative content, and were more distressed by them.

**Sense of coherence**
Sense of coherence refers to a personal orientation towards life, ‘a way of seeing the world’, which is presumed to engender and enhance one’s health experience (Antonovsky, 1993). The Sense of Coherence scale (SOC; Antonovsky, 1987) has been translated into at least 33 languages and measures three factors: comprehensibility (e.g. ‘Do you have very mixed-up feelings and ideas?’), manageability (e.g. ‘Do you have the feeling that you’re being treated unfairly?’), and meaningfulness (e.g. ‘How often do you have the feeling that there’s little meaning in the things you do in your daily life?’) (Monzi, 1998). Bengtsson-Tops & Hansson (2001) found that sense of coherence was related to mastery, self-esteem and social support, but was negatively associated with psychopathology. During an 18-month follow-up period, changes in sense of coherence were positively correlated with changes in subjective quality of life, general health, global well-being and psychosocial functioning.

**Recovery**
Service users and service user organisations have been active in proposing the assessment of outcomes beyond symptoms. They argue that psychological/social recovery is possible even with positive and negative symptoms, particularly when a person learns how to manage ongoing symptoms and relapses. They want to ‘make the important measurable, not the measurable important’ (Roberts & Wolfson, 2004). Central to the concept of recovery is hope and leading a meaningful life.

A number of measures to assess recovery have emerged. The Mental Health Recovery Measure (MHRM; Young & Ensing, 1999) was designed following interviews and focus groups with 18 consumers. It assesses three phases of recovery: overcoming sickness; discovering and fostering self-empowerment; and striving to attain overall well-being and reach new potentials. The Recovery Assessment Scale (RAS; Corrigan et al., 1999) is based on the narratives of service users and assesses personal confidence and hope, willingness to ask for help, goal and success orientation, reliance on others and symptom coping. Finally, Andersen et al. (2006) developed the Stages of Recovery Instrument (STORI), a 50-item measure covering five stages: moratorium, awareness, preparation, rebuilding and growth. Given the relatively lengthy recovery measures, they note that ‘a single, relatively short measure capturing this complex construct would prove invaluable’. Recovery has been found to be inversely related to symptoms and positively correlated with quality of life and empowerment (Corrigan et al., 1999).
PREDICTIVE VALIDITY OF SUBJECTIVE ASSESSMENT OF TREATMENT

Alongside increasing public and political interest in patient-reported outcomes, there is a scientific basis for their assessment in evaluating illness and treatment. This stems from the role of certain subjective criteria in predicting treatment outcome. In particular, patients’ ratings of satisfaction with care and the quality of the therapeutic relationship have been found to predict outcome of the treatment of schizophrenia across in- and out-patient settings. People with a more positive assessment of treatment and the therapeutic relationship tend to have a more favourable outcome (Frank & Gunderson, 1990; Pribe & Gruyters, 1995; Tattan & Tarrier, 2000), including reduced symptom severity, reduced rehospitalisation, and improved quality of life and social functioning. It may be that a better therapeutic relationship leads to greater acceptance of treatment, which in turn leads to a better outcome (Chue, 2006).

OVERLAP BETWEEN SUBJECTIVE EVALUATION CRITERIA

In studies that have assessed a range of patient-reported outcomes, the subjective indicators have regularly been found to be correlated. Such correlations occur even when there is no direct item overlap. It has been suggested that there is one factor, reflecting a general tendency for more positive or negative appraisal of one’s life situation, which explains more than half of the variance of constructs such as subjective quality of life, self-rated needs for care, self-rated symptoms and treatment satisfaction. Such a factor has been identified in cross-sectional and longitudinal analyses. It reflects a general appraisal tendency, and is strongly associated with patients’ mood (Pribe et al, 1998; Fakhoury et al, 2002). A small number of items were sufficient to capture this general factor (Fakhoury & Pribe, 2002; Hansson et al, 2007). Only changes in treatment satisfaction from baseline to follow-up loaded on a separate factor (Hansson et al, 2007). Similar empirical analyses have not yet been conducted on the constructs evaluating the self (self-esteem, empowerment, etc.). However, the correlations reported to date suggest that a large common factor may also account for much of the variance across these constructs.

Thus, there seems to be a mood-dominated general tendency for more or less positive subjective ratings of one’s situation, condition and external events. New scales may be developed in the future which are based on a new concept of subjective outcomes but still capture patient ratings and appraisals. It can be assumed that the results from such scales will also be influenced by the same general tendency. The challenge for further empirical work might be to find a simplified method for assessing the general tendency for more or less positive ratings and to capture the remaining construct-specific variance with scales whose development is closely linked to theoretical models for each construct (Hansson et al, 2007).

OUTCOME MANAGEMENT

In mental healthcare there is considerable interest in the potential for outcome management to improve clinical performance and patient care. Outcome management has been defined as a ‘technology of patient experience designed to help patients, payers and providers make rational medical care-related choices based on better insight into the effect of these choices on the patient’s life’ (Ellwood, 1988: p. 1551). Four factors characterise outcome management: greater use of standards and guidelines; routine assessment of patient functioning at appropriate time intervals; pooling outcome data on a massive scale; and dissemination of these results to relevant decision makers. The ultimate aim of outcome management is to improve clinical performance and patient outcomes (Smith et al, 1997).

Slade et al (2006) conducted a randomised controlled trial of routine assessment of patient-reported outcomes. Monthly postal questionnaires assessing needs, the severity of mental health problems and the therapeutic alliance were completed by patients (and staff). Patients also completed a quality of life assessment. The feedback was provided by post to patients and staff. The intervention did not improve patient-reported outcomes but reduced psychiatric in-patient days. Outcomes were rated by patients outside clinical consultations and the results later made available to clinicians. Such an approach makes it difficult to know whether the outcome management process ever had an impact on how clinicians really managed their patients. Incorporating the assessment and feedback of outcomes into routine clinical encounters so that it directly affects clinical management might make the process more meaningful for both clinicians and patients, and be more likely to improve patient outcomes.

Prieb et al (2007) conducted a randomised controlled trial of routine assessment and feedback of patient-reported outcomes in six European countries. They devised a computer-mediated intervention to structure patient–clinician dialogue (DIALOG) to be consistently patient-centred and to simultaneously assess patient outcomes. The outcomes were quality of life, needs for care, and satisfaction with treatment, all based on the patient’s view at the time of the dialogue.

This procedure was designed to alter interactions so that the patient’s views on their situation and needs for care were the central point of treatment discussions and the patient’s view on what kind of help would improve their situation was made explicit. Patients’ responses were entered onto a handheld computer/ laptop and fed back immediately in tabular and graphical screen displays for the patient and clinician to discuss together. The underlying rationale was that providing patients and clinicians with this information would lead to explicit negotiation about what the patient wanted and what the clinician could do about it. This in turn would improve subsequent care and the patient’s quality of life. The intervention reduced unmet needs and increased both treatment satisfaction and subjective quality of life over a 1-year period.

CONCLUSIONS

On the basis of the evidence, incorporating subjective assessments into the treatment process itself and in treatment evaluation is warranted. Routine assessment of patient-reported outcomes (e.g. needs for care) might reduce potential discrepancy between patients and professionals, and improve outcome, although there is little empirical evidence to support this assumption at present. Structuring patient–professional communication around the patient’s view of their medication, needs or quality of life (or a combination) prior to or during the consultation might have a positive impact on treatment provision and outcome.
In treatment evaluation, some subjective assessment of treatment is indicated because it predicts treatment outcome and it reflects the view of the 'consumer' of care. Although different constructs might appear to be conceptually distinct, there is significant empirical overlap between needs, assessment of treatment and the therapeutic relationship. This overlap reflects a general tendency, largely influenced by mood, to view one's situation positively or negatively. Of the constructs evaluating treatment, satisfaction with treatment is the outcome with the greatest evidence base, with more than 30 years of research.

The constructs that focus on self-evaluation are also overlapping. Empowerment, self-esteem and recovery are all interrelated, as are self-esteem and sense of coherence. Greater empowerment, self-esteem, sense of coherence and recovery are, in turn, associated with enhanced quality of life.

In conclusion, although the constructs might appear to be conceptually distinct, there is substantial empirical overlap. Developing a conceptual framework to classify existing and emerging patient-reported outcomes might be helpful to clarifying the specific contribution of each. Despite the intuitive appeal of proposed new patient-reported outcomes, they should not be developed unless they can be shown to be sufficiently independent of existing outcomes. More than one such outcome should not be used in a single study unless there is a specific hypothesis to justify this. For the future development of patient-reported outcomes, the most pressing issues are conceptual clarity and the consideration of empirical findings, particularly the association with other outcomes.

REFERENCES


Ruggieri, M. & Dall’Agnoila, R. (1993) The development and use of the Verona Expectations for Care Scale (VECS) and the Verona Service Satisfaction Scale (VSSS) for measuring expectations and satisfaction with community-based psychiatric services in patients, relatives and professionals. Psychological Medicine, 23, 511–523.


Outcomes of public concern in schizophrenia

IAIN KOOYMAN, KIMBERLIE DEAN, SAMUEL HARVEY and ELIZABETH WALSH

Background Schizophrenia is known to be associated with a range of adverse outcomes, which have an impact at the societal level and are therefore of public concern.

Aims To examine the epidemiology and methods for measuring six adverse outcomes in schizophrenia: violence, victimisation, suicide/self-harm, substance use, homelessness and unemployment.

Method A review of the literature was carried out for each adverse outcome, with attention to critical appraisal of existing measurement tools.

Results Schizophrenia is associated strongly with all six outcomes, although research has mainly focused on violence. Each outcome acts as a risk factor for at least some of the other outcomes. There are few standardised or validated measures for these ‘hard’ outcomes. Each measure has inherent biases but a growing trend is for these to be minimised by using multiple measures.

Conclusions A single instrument which systematically measures multiple societal outcomes of schizophrenia would be extremely useful for both clinical and research purposes.

Declaration of interest None.

The asylum movement of the 19th century could be regarded as part of a state-guided sanitary movement to cleanse society of the harmful impact of those with mental illness. Although stigmatisation and fear were instrumental in this process, there is now strong evidence for genuine adverse outcomes of schizophrenia on society. Deinstitutionalisation and community care, which have become widespread since the 1970s, have re-exposed the general public to such outcomes, accompanied by a fear of violence, and particularly homicide perpetrated by people with schizophrenia, fuelled by media attention. Suicide and self-harm are much more prevalent outcomes in this group, however, and victimisation of people with schizophrenia is especially neglected. Substance misuse, unemployment and homelessness are also prevalent outcomes of public concern. The prevalence and risk factors for each of these six adverse outcomes in schizophrenia will be reviewed, with an examination of the contribution to society as a whole. There are few validated instruments for measuring these ‘societal outcomes’, but their assessment at both the individual and population level will be considered.

VIOLENCE

Prevalence and risk factors

It is now widely accepted that people with schizophrenia are more likely to behave violently. Varying estimates of the prevalence and relative risk of violence in schizophrenia are dependent on the definition of schizophrenia, the type of violence measured and the location of the study. There is also no consensus as to which variables should be treated as confounding factors or mediators. Unselected birth cohorts have reported relative risks of between 2 and 7 times for serious violence compared with the general population (Tiihonen et al., 1997; Arsenault et al., 2000; Brennan et al., 2000). People with schizophrenia have been shown to be convicted of a greater number of violent crimes than their neighbours of a similar age (Wallace et al., 2004) and schizophrenia is overrepresented in prisoners (Teplin, 1990; Eronen et al., 1996). Although schizophrenia independently increases the risk of committing violence (Brennan et al., 2000), this risk is increased significantly by comorbid substance misuse (Wallace et al., 2004), personality disorder (Moran & Hodgins, 2004), a lack of adherence to medication (Swanson et al., 1997) and acute psychotic symptoms (Taylor, 1998).

Risk to society

With most research to date focusing on relative risk, it is encouraging to see estimates of absolute risk emerging in the literature. The population attributable risk (i.e. the fall in levels of violence in society that would occur if violent incidents by people with schizophrenia were discarded) is an approximate calculation. This approach assumes causality between schizophrenia and violent conviction and fails to take account of associated factors, such as substance misuse and personality disorders. Wallace et al. (2004) estimated that 6–11% of violent convictions are attributable to schizophrenia. Fazel & Grann (2006) found a population attributable risk of just 2.3%, which increased to 5% for psychosis. They suggest that in countries with more liberal gun laws, the attributable risk is lower for homicide, but others argue that those with schizophrenia are responsible for 5–10% of homicides irrespective of the baseline homicide rate (Wallace et al., 2004).

Measurement

Measurement of violent behaviour has relied upon various single or combined sources of information (self-report, informant, case notes, official records). All single sources bias towards underreporting: self-report from a desire for social acceptability or fear of adverse consequences of reporting; informants, often nominated by patients, being unreliable or unaware; and case notes being invariably incomplete. The proportion of violent acts leading to arrest, prosecution and conviction varies with the intensity and quality of policing, the behaviour of the suspect, the availability of diversion to the mental health system and the severity of offence. Most people who are violent are not convicted (Elliott et al., 1986). Only the more serious violent
acts lead to conviction; hence the association between schizophrenia and more minor forms of violence is impossible to estimate from official sources.

The recent use of multiple combined measures has improved the detection of violent behaviour. Steadman et al (1998) showed that the detection of violence increased steadily as methods were combined, and reached six times the rate of official convictions alone. Multiple measures require judgements about what constitutes a single violent event and handling inconsistencies between reports.

The definition of violence varies enormously between studies, and most neglect contextual aspects. The MacArthur Community Violence Interview (Steadman et al, 1998) in the USA is an important step towards consistency. It measures lifetime violence, and includes information on recent aggressive behaviour and victimisation. It incorporates a clear and structured definition of different levels of violence and considers the context for each episode. There is also a version for use with collateral sources. Encouragingly, its use is increasing (Elbogen et al, 2006; Swanson et al, 2006).

**Predicting violence**

Measuring violence is less problematic than predicting it. Assessing the risk of violence has become an increasingly important part of clinical practice in psychiatry, with time and resource implications. The clinical usefulness of specific risk assessment procedures depends on: (a) the accuracy of prediction (predictive validity); (b) the applicability to the patient group; and (c) the ability of clinicians to act on the results to reduce predicted risk.

Predictive validity has been at the heart of the debate concerning two differing approaches – actuarial v. clinical risk assessment. The former relies on the identification of largely static risk factors defining at-risk groups within populations while the latter is an individually focused case formulation, which underpins routine clinical practice. To combine the advantages and minimise the disadvantages of the two approaches, several structured risk assessment instruments have been devised and tested (Dolan & Doyle, 2000), including the Violence Risk Scale (VRS; Wong & Gordon, 2000).

A statistical assessment of predictive validity is essential both for considering the clinical value of a particular instrument and for comparing instruments. Receiver operating characteristics (ROC) analysis integrates the concepts of sensitivity and specificity, and are relatively independent of the base rate of violence within the population (Kroner, 2005). A recent UK study compared the relative efficacy of the Historical Clinical Risk 20 items scale (HCR-20; Douglas et al, 2001), the Psychopathy Checklist Screening Version (PCL-SV; Hart, et al, 1995) and the Offender Group Reconviction Scale (OGRS; Copas & Marshall, 1998) prospectively over 2 years in a group discharged from a medium secure unit (Gray et al, 2004). All three instruments were predictive of offending over the follow-up period, but the purely criminogenic scale (OGRS) performed best. This finding that actuarial instruments outperform even structured clinical assessments in mentally disordered offenders is consistent across different settings (Bonta et al, 1998), but both types of assessment outperform unaided clinical judgement. However, instruments validated in offenders may have less predictive validity in general adult than forensic psychiatry. The HCR–20 has been validated in both settings (Douglas et al, 2001).

In clinical practice the usefulness of any risk assessment method will also depend on the implications for intervention. Static factors such as gender and past criminal behaviour offer limited scope to inform clinical intervention. Consideration of dynamic, clinical factors, such as active psychotic symptoms and substance misuse, may contribute more to the usefulness of a risk assessment instrument in clinical practice (Mills, 2005), enabling the shift from risk assessment to risk management or risk reduction.

Imperfect risk prediction has serious implications for individuals. Even instruments with relatively high predictive validity will generate both false-positives and false-negatives. The potential implications have been elegantly demonstrated by Buchanan & Leese (2001) who pooled results from 23 studies employing violence risk assessments and concluded that 6 people would need to be detained to prevent one violent act. Routine violence risk assessment might also detract from the consideration of other outcomes, such as those reviewed below.

**VICTIMISATION**

**Prevalence and risk factors**

People with severe mental illnesses such as schizophrenia are more likely to be victims of violence than perpetrators of a violent act (Brekke et al, 2001). Silver (2002) reported that people with severe mental illness and/or personality disorder were more than twice as likely to be the victims of violence than their neighbours. Recent US figures are much higher (Teplin et al, 2005) and are supported by findings from the Dunedin Study, in which over half of those with schizophreniform disorder reported being assaulted in a 12-month period (Silver et al, 2005).

It has been suggested that this increased risk of victimisation arises from increased aggressive behaviour. Although this may play a part, the increased risk of victimisation in people with psychosis remains irrespective of the individual’s own violent behaviour (Hiday et al, 2002; Silver, 2002). People with schizophrenia now live within the community and Silver (2002) has shown that their victimisation can be mediated by conflict within social relationships. Elevated rates have also been found to be prospectively associated with comorbid personality disorder, young age at illness onset, previous victimisation and infrequent contact with family members (further details available from K.D.).

**Risk to society**

Little is known about the impact of victimisation on either the individual or society. It is likely that victims of violence who have schizophrenia will be particularly vulnerable to a range of adverse outcomes, such as homelessness (Lam & Rosenheck, 1998), which have significant cost implications.

**Measurement**

Victimisation is poorly recognised in clinical practice (Cascardi et al, 1996), often neglected in schizophrenia research and optimal methods of measurement have yet to be established. Two types of instruments have been used. Questionnaires have been designed for use with people with mental disorders, but not specifically to examine victimisation. The MacArthur Community Violence Interview includes a number of questions on victimisation and its context (Silver, 2002). The Lancashire Quality of Life Profile includes items on experience of victimisation, but without detail of the frequency, severity or context (Oliver, 1991). Questionnaires have also been designed to examine victimisation in the general population. The National Crime Victimisation Survey was applied to a sample of people with serious mental illnesses by Teplin et al, 2005 who
described the instrument as the most comprehensive available to assess victimisation because it elicits detailed information about each event reported. The instrument required some modification for use with people with mental disorders.

As many acts of violence are not reported to the police (and this may be more likely for victims with mental illnesses) self-report measures will continue to be the best method for obtaining data on victimisation. Reporting past victimisation may be subject to recall difficulties and may not be reliable. Incorporation of ‘bounding interviews’ to establish reference points for future recalling of index events might reduce ‘telescoping’, whereby incidents occurring prior to the required recall period are reported (Teplin et al., 2003). Collateral sources (family members, keyworkers or residential support staff), although generally likely to underestimate victimisation, may complement participant-reporting and enable some assessment of reliability. As with the measurement of all societal outcomes, the use of multiple sources of information is optimal.

Attention has been focused on establishing the prevalence of victimisation and associated risk factors, rather than understanding in detail its nature, context and impact on those with schizophrenia and other serious mental disorders. Future measures of victimisation should consider factors such as acute symptoms, service contacts and presence of comorbid illness. In addition to exploiting multiple sources, instruments should be specifically designed for people with mental illness and should assess victimisation in detail.

SUICIDE AND SELF HARM
Prevalence and risk factors
Suicide is a significant cause of premature death in people with schizophrenia (Caldwell & Gottesman, 1992), with lifetime estimates ranging from 5 to 13% (Miles, 1977; Caldwell & Gottesman, 1990; Palmer et al., 2003). Most suicides occur soon after illness onset (Palmer et al., 2003) and may have increased greatly over the past century (Healy et al., 2006). Non-fatal acts of self-harm are also increased, with a study of people with chronic schizophrenia finding that 38% had at least one episode of self-harm in a 2- to 12-year follow-up period (Breier et al., 1991).

A recent meta-analysis identified the following as risk factors for suicide in schizophrenia: recent loss; fear of mental disintegration; agitation or motor restlessness; poor adherence to treatment; drug misuse; and previous depressive disorders and suicide attempts (Hawton et al., 2005). Suicidal behaviour in individuals with schizophrenia does not appear to be associated with particular psychotic symptoms. The usual higher incidence of self-harm in females is not present in schizophrenia (Hawton et al., 2003) and, strikingly, people with schizophrenia from more affluent socio-economic groups are at increased risk of self-harm (further details available from the authors). Approximately 20% of suicides in those under 35 are accounted for by schizophrenia (Appleby et al., 1999a).

Measurement
Accurate estimation of suicide rates is difficult; official statistics and coroners’ reports are known to underestimate suicide rates, but such errors do not invalidate epidemiological conclusions based on these figures (Sainsbury & Jenkins, 1982; Speechley & Stavraky, 1991). Some estimates rely on proportionate mortality (the percentage of those dead who died by suicide) rather than case fatality rates (the percentage of a sample of patients who will die by suicide). The use of proportionate mortality rates assumes a constant rate of suicide, which given the increased rate of early suicide in schizophrenia might lead to an overestimate of the lifetime suicide risk (Palmer et al., 2005).

A number of risk factors have been consistently associated with suicide in schizophrenia, but their low sensitivity and specificity, plus the rarity of suicide, diminish their clinical usefulness. Evaluating the predictive power of suicide risk factors in psychiatric in-patients, Powell et al. (2000) found several to be strongly associated, but the resulting model was unable to predict the majority of suicides without an unacceptably high false-positive rate.

The definition of self-harm is not well established (Skegg, 2005). Behaviours vary and there is no consensus on inclusion of suicidal intent, which can be difficult to measure in psychosis. Clinical records underestimate self-harm compared with self-report questionnaires (Hawton et al., 2002), but self-report alone may be unreliable. Some studies combine self-report with review of routine case records. Instruments including a limited number of items relating to self-harm have been used to estimate its prevalence in schizophrenia. These include the WHO Life Chart (World Health Organization, 1992), the Structured Clinical Interview for DSM (SCID; Spitzer et al., 1994), the Functional Assessment Rating Scale (FARS; Ward & Dow, 1995) and the Psychiatric and Personal History Schedule (PPHS; Jablensky et al., 1992).

The European Parasuicide Study Interview Schedule (EPIS) has been specifically developed to examine parasuicidal behaviour, suicidal thoughts and associated factors in detail (Platt et al., 1992), but has only been used to a limited extent in samples with psychotic disorders (Nordentoft et al., 2002).

SUBSTANCE USE
Prevalence and risk factors
In the USA 40–60% of people with schizophrenia misuse substances, excluding cigarettes (Cantor-Graae et al., 2001). The pattern of substances misused varies locally but rates are universally higher than in the healthy population (McCreddie et al., 2002). Substances misused include all substance classes and appear to be increasing dramatically (Boutros et al., 1998), although proportionally to the rise within the general population (Wallace et al., 2004).

Substance misuse is increased prior to the onset of schizophrenia. This might be due to causality of psychosis by drugs such as cannabis (Arsenault et al., 2004; Fergusson et al., 2003) or confounders such as a shared underlying neurological vulnerability (Janowsky et al., 1973; Liberman et al., 1986) or antisocial personality disorder (Reiger et al., 1990). Substance misuse is also an outcome of schizophrenia. A substantial number of people use drugs for the first time after the onset of schizophrenia (Hambahrt & Hafner, 1996). Such patients with dual diagnosis report using street drugs to counter depression and anxiety (Dixon et al., 1990; Addington & Duchak, 1997), negative symptoms such as apathy and anhedonia (Pristach & Smith, 1996), and to assist sleeping and reduce extrapyramidal side-effects. Cocaine use may temporarily reduce negative symptoms (Serper et al., 1996). Evidence that people use street drugs to treat positive symptoms is equivocal. People with schizophrenia often feel alienated from society (Sainsbury Centre for Mental Health, 1998) and, rejected by peers, may drift into networks of drug users, who may be more accepting of them (Lamb, 1982).

Substance misuse is clearly an adverse outcome: people with dual diagnosis are
generally younger, less adherent to treatment (Swolford et al, 1996), have more positive symptoms (Hambrecht & Hafner, 1996), more psychiatric admissions (Hunt et al, 2002), higher rates of violence (Hodgins, 1992; Scott et al, 1998), are more likely to die by suicide (Appleby et al, 1999b), be unemployed (Seibyl et al, 1993), homeless (Drake et al, 1991; Soyka et al, 1993) and create excess service costs (Hoff & Rosenheck, 1999). The extent of the damage is underlined by this group's superior premorbid intellectual functioning and socio-economic status compared with people with schizophrenia who do not misuse substances (Kirkpatrick et al, 1996; Sevy et al, 2001). Much of the three-fold higher mortality in schizophrenia can be attributed to excess substance misuse, especially cigarette smoking (Brown et al, 2000).

**Measurement**

Clinicians and family informants are poor at estimating substance misuse in the absence of dependency, and patients grossly underreport their use (particularly for stimulants and opiates) when compared with toxicology screens (Swartz et al, 2003). Detection by professionals depends on the level of training in drug/alcohol issues and familiarity with the patient (Ananth et al, 1989). Staff suspicion and questioning should be combined with toxicology screens, but these also require staff training and provide only binary outcomes (i.e. used/not used). Saliva tests avoid the risk of patients corrupting samples and awkward supervision, but it remains unclear whether they are more or less accurate than urine tests. Breathaliser tests are practical and valid for measuring alcohol intoxication. For detecting more distal substance use, radioimmunoassay of hair specimens is non-intrusive and reliable (Swartz et al, 2003).

‘Use’ can be quantified by frequency, quantity or duration, and should be differentiated from ‘misuse’ and ‘dependency’, but for convenience, poorly defined pooled categories have been preferred. Common examples include ‘substance use disorder’ (Mueser & Drake, 1998) and ‘problem use’ which has been variably equated to harmful or dependent use combined (McCreadie, 2002), or any use (for example Duke et al, 2001). Studies vary in the extent of substance inclusion, particularly of legal (nicotine, caffeine, alcohol) and prescribed substances (benzodiazepines, anticholinergics). Substance use diagnoses can refer to current, past or lifetime criteria.

Most research studies use case notes or unstructured interviews. Structured interviews minimise information variance and are more reliable (Blanchard & Brown, 1998). Some standardised measurement tools are listed in Table 1 but these are rarely used outside of research. Multiple measures are increasingly being used (Swartz et al, 2006).

Routine screening for substance misuse in people with schizophrenia is an important component of assessing risk and planning treatment. Self-report measures assessing readiness to change are reliable (Carey et al, 2001). However, evidence for effectiveness of psychological interventions targeting substance misuse over standard care for people with schizophrenia has been lacking (Ley et al, 2000), but is improving (Haddock et al, 2003).

### HOMELESSNESS

**Prevalence and risk factors**

Homelessness is a well recognised outcome of schizophrenia but there have been few attempts to quantify it. Rates vary across borders and time. A US community study (Folsom et al, 2005) found that about a fifth of more than 4000 people with schizophrenia had no fixed address, which was 2.4 times higher than for major depression. The European Schizophrenia Cohort...
Table 2  Studies of the effectiveness of individual placement schemes for people with severe mental illness

<table>
<thead>
<tr>
<th>Reference</th>
<th>Outcome difference for those in employment, %</th>
<th>Other employment outcomes measured</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IPS</td>
<td>Controls</td>
</tr>
<tr>
<td>Drake et al (1996)</td>
<td>78.1</td>
<td>40.3</td>
</tr>
<tr>
<td>Drake et al (1999)</td>
<td>60.8</td>
<td>9.2</td>
</tr>
<tr>
<td>Lehman et al (2002)</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td>Mueser et al (2004)</td>
<td>73.9</td>
<td>18.2/27.5</td>
</tr>
<tr>
<td>Gold et al (2006)</td>
<td>64</td>
<td>26</td>
</tr>
<tr>
<td>Latimer et al (2006)</td>
<td>47</td>
<td>18</td>
</tr>
</tbody>
</table>

IPS, individual placement scheme.
1. 18.2% for psychosocial rehabilitation and 27.5% for standard care.

(Bebbington et al, 2005) found that 32.8% of the British sample had experienced homelessness in their lifetime compared with 8.4% in Germany and 12.9% in France. The rate in London was even higher (43%) and 13.2% of the British sample had experienced rootlessness, despite those who were currently roofless being excluded from the study.

Large US schizophrenia studies consistently find homelessness to be associated with substance misuse and severity of symptoms, but have also found associations with African–American ethnicity (Folsom et al, 2005), lower global functioning (Olsson et al, 1999) and more autistic preoccupations (Opler et al, 2001).

Housing instability in people with schizophrenia predisposes to institutionalisation in prisons and hospitals (Appleby & Desai, 1987), non-adherence with treatment, psychosocial problems (Drake et al, 1989) and decreased quality of life (Lehman et al, 1995). Physical and sexual abuse are extremely common in both male and female homelessness (Wenzel et al, 2000). Mortality is more than 3 times higher in the homeless (Hibbs et al, 1994). Outcomes may be poorer for homeless people with schizophrenia in urban compared with rural areas (Drake et al, 1991).

Risk to society
The proportions of homeless people with schizophrenia vary with levels of social and mental health provision, for example 12% for males in Munich (Fichter et al, 1999) and 23% for males in Sydney (Teesson et al, 2004), but are higher in urban areas and significantly higher in the female homeless (46% in Sydney; Teesson et al, 2004).

Measurement
‘Rooflessness’ refers to those living on the streets, and defines the group of most public concern but which is hardest to locate or follow-up. Most studies (e.g. Folsom et al, 2005) use a looser definition of having no fixed address and include people living in hostels and emergency accommodation. Some researchers have further widened the concept to include a spectrum of ‘housing instability’, signifying tenuousness of housing tenure and associated stress (Drake et al, 1991). This group of so-called sofa-surfers move frequently between friends, family and emergency housing.

There are no valid national databases of housing because of unofficial rental, unregistered housing by friends and family, and the rapid movements of individuals. Case manager rating scales of housing instability have been used, such as a 5-point Likert scale screening device (Drake et al, 1991), which rates accommodation from ‘highly supportive’ to ‘highly stressful’. This may help to identify people with housing problems who can then be given a more detailed structured interview.

However, people who are living on the streets, especially those with prominent negative symptoms or an itinerant lifestyle, are less likely to be in regular contact with mental health services, thus rates of homelessness in people with schizophrenia may be underestimated. Assertive screening of the homeless for mental illness might reduce the exclusion of this group from mental health services.

UNEMPLOYMENT
Prevalence and risk factors
The European Schizophrenia Cohort (Bebbington et al, 2005) found that only 11.5% of the British sample were actively employed, including sheltered employment. The French rate was similar (12.9%) but the German much higher (30.3%). Estimates of about 22% have been made in both the USA (Mechanic et al, 2002) and Australia (Carr et al, 2004). More encouragingly, the International Study of Schizophrenia (IsoS) found that 37% of people with schizophrenia had received paid work for most of the past 2 years (Harrison et al, 2001), but attrition rates were high.

The gradual decline in rates of employment over many years leading up to diagnosis in a large Danish population cohort (Agerbo et al, 2004) suggests impairment during the prodromal phase. Rates of employment deteriorate further after first presentation (Mechanic et al, 2002; Agerbo et al, 2004). Among people with schizophrenia, past admission to hospital predicts
current unemployment (Munk-Jørgensen & Mortensen, 1992).

Unemployment is associated with decreased quality of life in schizophrenia (Caron et al., 2005). Lewine (2005) showed that job expectation prior to the onset of schizophrenia significantly correlated with depression and hopelessness, and both were increased in higher socio-economic groups.

Educational attainment is the best protective factor for employment in people with schizophrenia, as in the general population (Mechanic et al., 2002). Cognitive functioning is a significant predictor of job tenure (Gold et al., 2002) and response to vocational rehabilitation (McGurk & Mueser, 2004).

Risk to society

The cost of unemployment owing to schizophrenia is considerable. Numbers of American recipients of disability benefits for schizophrenia rose by 35% between 1994 and 2003 (Rosenheck, 2006). Not surprisingly several initiatives are underway to improve employment in this group. A Cochrane review (Crowther et al., 2001) concluded that supported employment, such as individual placement and support (Bond et al., 1997), is more effective than pre-vocational training for obtaining competitive employment.

Measurement

Employment is not an all-or-nothing phenomenon and should be considered in terms of quantity and quality, both for the individual and research purposes. Studies examining the impact of individual placement schemes in assisting attaining employment have used quite consistent measures of employment (Table 2). These studies all principally examined the proportion of people with mental health problems who attained competitive employment, which has been defined as a job in which payment is at least the minimum wage, is not reserved for people with disabilities and fewer than half of the person’s co-workers have disabilities (Latimer et al., 2006).

However, sheltered employment, although less lucrative and unable to supply the same level of integration, can also increase skills and self-esteem.

Quantity of employment can be measured by either hours worked or income earned. Quality of employment can be measured by: (a) job tenure, i.e. how long each job is held; working for short periods in a variety of jobs is likely to be less fulfilling and give a lower sense of financial security; (b) job satisfaction; and (c) secondary benefits, such as social contact, quality of life, etc.

The studies described above all used a combination of self-report and keyworker ratings every 6 months. Latimer et al. (2006) supplemented these interviews with two monthly telephone interviews. However, self-report measures may overestimate levels of employment owing to bias arising from social desirability, denial and grandiosity. Few studies have included employer interviews, which participants may not consent to.

Receiving benefits has been used as a proxy measure for employment. However, some unemployed people are supported by savings or family members/partners and are either ineligible or choose not to collect benefits. Another group collect benefits but work legally part-time, or work occasionally or frequently ‘off the books’.

Conclusions

Schizophrenia is strongly associated with a range of adverse outcomes, which have an impact at the societal level. There is much intercorrelation between these outcomes, suggesting the possibility of a domino-like effect for an individual person, whereby each outcome leads to another. To limit this downward spiral, it is crucial that all of these outcomes are considered simultaneously. Reducing these outcomes would require implementation of a combination of strategies at national, local and patient levels (e.g. the matrix model of healthcare provision; Tansella & Thornicroft, 1998).

There are few standardised definitions, let alone validated measures for these outcomes, which makes comparison or collation of research findings problematic. A systematic review of studies proposing implementation of routine mental health outcome measures (Slade, 2002) identified few studies examining any of the outcomes discussed here. For clinical purposes, therefore, the mere consideration of these outcomes, alongside thorough assessment and the use of multiple information sources, allows the best chance of a positive outcome. For research purposes, a collection of validated and brief assessments or even a single instrument to systematically measure these societal outcomes would be extremely useful.

References


Comorbid non-alcohol substance misuse among

A randomized clinical trial of supported

South

et al

et al

et al

Substance abuse in schizophrenia: a review of the

531^533

(1996)

people with schizophrenia. Epidemiological study in

Drake, R. E., McHugo, G. J., Becker, D. R.,

Diagnosis of alcohol use disorders in schizophrenia.


Carr, V. J., Lewin, T. J., Neil, A. L.,


Elliott, D. A., Huizinga, D. & Morse, B. J. (1986)


Early reactions to cannabis predict later dependence. Archives of General Psychiatry. 60, 1033–1039.


Mental illness in homeless men and women in Munich. Psychiatrische Praxis. 26, 76–84.


The Severity of Dependency Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction. 90, 607–614.


Substance abuse and the onset of schizophrenia. Biological Psychiatry. 40, 1155–1163.


Hodgins, S. (1993)


The cost of treating substance abuse patients with and without comorbid psychiatric disorders. Psychiatric Services, 50, 1309–1315.


Medication compliance and comorbid substance misuse in schizophrenia: impact on community survival 4 years after relapse. Schizophrenia Research. 54, 253–264.


Effects of homelessness on the quality of life of persons with severe mental illness. Psychiatric Services, 46, 922–926.


Social class of origin, lost potential, and hopelessness in schizophrenia. Schizophrenia Research. 76, 329–335.


Treatment programmes for people with both severe mental illness and substance misuse. Cochrane Database. Cochrane Database Issue 2. Update software.


Sainsbury Centre for Mental Health (1999) Keys to Engagement: Re-View of Care for People with Severe Mental Illness who are Hard to Engage with Services. Sainsbury Centre for Mental Health.


Hospitalisation as an outcome measure in schizophrenia

TOM BURNS

Background People with schizophrenia comprise the majority of patients with severe mental illness recruited to recent mental health service studies of new teams (e.g. assertive outreach, crisis resolution). Reduction in hospitalisation has been the most consistent outcome measure in these studies, but results are inconsistent.

Aims To understand inconsistency of results from studies using hospitalisation as an outcome measure.

Method The advantages and disadvantages of hospitalisation are explored, including the ways in which it is recorded. Regional variation in outcomes and the impact of control services are reviewed.

Results Hospitalisation has face validity as an outcome but translates poorly between differing healthcare contexts. These variations can be exploited positively to distinguish potentially effective ingredients in community care (outreach, combined health and social care, team structure) from redundant components.

Conclusions Hospitalisation is a good proxy outcome measure in schizophrenia care in randomised controlled trials, but the dangers of extrapolating to new contexts require care.

Declaration of interest T.B. has received payments for lectures and consultancies from Eli Lilly, Janssen and Otsuka in the past 5 years.

The past 25 years have witnessed an explosion in mental health services research. A 1980 review of research in community mental health services (Braun et al, 1981) cited a dozen studies and concluded that there was little evidence that the newer services sustained people longer outside hospitals. Stein & Test’s landmark study of assertive community treatment (ACT) was particularly influential because not only did it demonstrate reduced hospitalisation along with improved clinical outcomes (Stein & Test, 1980), but the accompanying paper indicated that it could achieve this without increased costs (Weisbrod et al, 1980). Not surprisingly these two findings stimulated an enormous interest in developing and evaluating such programmes. When Mueser et al (1998) reviewed the area they were able to cite 75 good-quality studies, and when Catty et al (2002) did the same 4 years later they had over 90 studies to draw on. The vast majority of these studies focused on those with severe mental illness and invariably the samples consisted mainly of people with schizophrenia – over 80% in the UK700 trial and the Department of Veterans Affairs study (Burns et al, 1999; Rosenheck et al, 1995).

HOSPITALISATION AS THE COMMON OUTCOME MEASURE

Not surprisingly in evolving mental health services there is a considerable range of outcome measures used; these reflect both the evolution of measures and the importance placed on these various outcomes. For some researchers symptom control might have been the goal, for some social functioning and community stability, for others quality of life or risk reduction. Hospitalisation is, however, overwhelmingly the most consistent outcome reported. Hospitalisation has been assumed to be a proxy for relapse in schizophrenia in the absence of a consensus on a clinically meaningful alternative measure. The use of pre-agreed changes in symptom scores, for example a 20% decrease or reduction to an agreed level of the Positive and Negative Syndrome Scale (PANSS) score, as in trials of antipsychotics, has not found favour with clinicians for service evaluations. They are rejected because of their sensitivity to prior levels of disturbance and insensitivity to key clinical features, such as self-neglect or hostility, that have a disproportionate influence on clinical management. As a consequence, hospitalisation has come to dominate randomised controlled trials (RCTs) of community interventions because it benefits from the assumed consistency of admission threshold in any local health system despite the known variation of these between systems.

Reporting of hospitalisation

Hospitalisation is generally reported in one of three forms in community studies.

Number of admissions

This is the simplest approach and consists of recording any psychiatric admission during the study period. The frequency of admissions is usually recorded during the follow-up period and outcomes reported in terms of admitted v. not admitted. This reporting has the advantage that it is immediately obvious to the reader, who may know little of the local circumstances or details about admission. If there are many patients with repeated admissions during the follow-up period then the mean number of admissions in the study categories may also be presented.

Time to admission

Time to readmission has been more used in relapse prevention studies than in community care studies. The difference between the timings of relapse in the experimental and control services are presented either as mean durations or, more usually, with survival curves (e.g. Kaplan–Meier).

Duration of in-patient care

The most common presentation of hospitalisation as outcome is by days of in-patient care within the agreed follow-up period. In schizophrenia trials hospitalisation data are rarely normally distributed and usually have a pronounced skew. The majority of patients usually have no admissions and a small number of patients account for most
of the in-patient days. Such data are best presented as medians rather than means, but planners prefer means so that they can calculate bed needs. It is increasingly common to assess bed-days with parametric statistics, presenting means, after subjecting the non-parametric results to bootstrapping techniques (Efron & Tibshirani, 1993). An advantage of duration of care is that it permits the pooling of hospitalisation data between studies with differing follow-up periods, because the durations can be recalculated as, for example, days per month or days per year.

ADVANTAGES OF HOSPITALISATION AS AN OUTCOME MEASURE

Face validity
The most obvious advantage of hospitalisation as an outcome measure is its face validity. All clinicians have a sense of when people with schizophrenia may need admission and what admission means for the patient, the family and the service. This understanding may, of course, be more illusory than real; the threshold for admission and the experience of admission may be very different in inner-city London and in a small town in Switzerland. However, a finding that an intervention halves admission rates or duration is immediately understandable and translatable to the clinician’s practice.

Utility
An understanding of changes in bed occupancy has direct utility for service planning. Indeed, it has been the translation of this outcome into projections of bed occupancy that has driven much of the research in this area and had an impact on service developments. There has been concern that the utility of research in this area has been exaggerated, either through naivety or in the service of economic imperatives. Well-recognised factors that inflate the effectiveness of newly established services (Coid, 1994), such as charismatic leaders, the recruitment of exceptional staff and the slow accrual of complex and resistant patients, have been ignored, leading to overoptimistic bed reductions.

Health economic analyses
Because hospitalisation is such a disproportionately expensive component of mental health services – still responsible for 80% of costs in many services (Leff et al., 2000) – careful recording of it is essential to any form of cost analysis. Mental health economic analyses require careful reading and careful interpretation. More than in any other branch of medicine the extent of the costing exercise is open to real debate – how much should housing and unemployment be included, how is informal care costed, etc.? Small, apparently unconnected, changes in living conditions can completely reverse the economic benefits of interventions (McCrone et al., 1994). Where studies include hospitalisation as an outcome such complications are unlikely, but conclusions about comparative costs within services require attention to local conditions. The difference between the costs of an in-patient day and an outpatient contact with a professional are not fixed. For example, the difference between the cost of an in-patient day and a case manager contact was much greater in Stein & Test’s study (Weisbrod et al., 1980) than in the UK700 study (Byford et al., 2000). Consequently how many case manager contacts would be paid for by a saved day in hospital would be very different in the two studies.

Despite these caveats, hospitalisation data are an essential component of health economic analysis and can make a powerful case for expanding or contracting different components in an integrated service. Careful costing of hospitalisation was responsible for dispelling the early myth that deinstitutionalisation was inevitably cheaper than hospital care and helped to identify levels of disability at which hospital care was cheaper overall (Knapp et al., 1990; Hallam et al., 1994).

DISADVANTAGES OF HOSPITALISATION AS AN OUTCOME MEASURE

‘Negative’ therapeutic goal
The most common criticism of hospitalisation as an outcome is a sense of its inadequacy in conveying normal and desirable clinical aims. Surely, its critics demand, there is more to psychiatry than simply keeping beds empty? Patient and family groups are often dismissive of a reduction in hospitalisation as evidence that services are more interested in an outcome relevant to themselves (i.e. reduced costs or administrative convenience) than to patients and families – improved well-being, quality of life and symptom control. This is a powerful criticism and not easily dismissed. Repeated attempts to contextualise hospitalisation as an outcome (explaining that it is a benchmark for clinical success, a proxy, rather than a direct measure) are necessary but often unsuccessful.

There are also disadvantages from a service development and delivery perspective of an exclusive focus on reduced hospitalisation. Sustaining mental health services relies on recruiting and retaining highly qualified staff, and for this the day-to-day business of care must be centred on the individual well-being of the patient directly in front of the staff member. Maintaining focus and motivation for the staff member and engaging the patient requires a clear therapeutic goal that can be shared and realised in that interaction (e.g. reducing distress, improving understanding of the illness or treatment, ensuring adherence to medication). Reducing bed occupancy is not one such shared goal. Reframing this as ‘promoting stability’ or ‘improving community tenure’ goes some way to presenting it as a desirable positive goal, but statistical probabilities are weak motivators in human behaviour. Clinical experience emphasises the need to identify the clinical practices and the interpersonal and patient-centred outcomes that lead to a goal of reduced hospitalisation (Wright et al., 2004) and enshrine these in operational policies (Burns & Finn, 2002).

Research distraction
Another criticism of hospitalisation as an outcome measure is that it can distract from efforts to explore the mechanisms of schizophrenia care. This criticism certainly does have salience in service development research (Burns et al., 1999), where preoccupation with organisation has led to a relative neglect of the operative components (Wright et al., 2004), but it is probably unwarranted in the area of schizophrenia outcomes. Current research in schizophrenia care demonstrates attention to a wide range of specified interventions, both pharmacological and psychosocial, and a wide range of outcome measures.

HOSPITALISATION AND RELAPSE

Independent assessment of relapse
Hospitalisation owes its current status as a research outcome principally to its assumed equivalence with relapse. Two recent developments question this legitimacy. First, the
increased thresholds for admission in hard-
pressed services or in tightly managed
services may require a specific degree of
severity of relapse for admission. Although
agreed definitions of relapse in pharmaceu-
tical trials have been long established based
on agreed changes (either absolute or per-
centage point changes) in symptom ratings
(such as PANSS score) they have rarely
been used outside drug trials. Where there
are regular ongoing assessments of clinical
status as part of a study it could be possible
to identify relapse independent of hospital-
isation. Several ongoing naturalistic and
observational studies, such as the Schizo-
phrenia Care and Assessment Programme
– UK (SCAP–UK; Burns et al, 2006), have
attempted to construct relapse criteria from
symptom changes, drug prescribing
changes and changes in contact frequency.
To date none of these attempts has been
replicated in published studies.

The Lambeth Early Onset study of early
intervention in psychosis has reported a re-
liable method for estimating relapse from
regular systematised assessments of case
notes (Bebbington et al, 2006). The
assessed relapses were strongly correlated
with independently assessed PANSS scores.
Whether such an approach will erode the
status of hospitalisation as an outcome
measure is as yet unclear. A series of studies
using such instruments might provide a guide
to the relationship between relapse rates and
hospitalisation rates in schizophrenia that
can then be used to scale up the inevitably
conservative hospitalisation rates.

Crisis resolution/home treatment
studies
Unlike research into case management or
assertive outreach, studies of crisis resolu-
tion/home treatment teams also use hospi-
talisation as a primary outcome measure
but without the assumption that a change
reflects a change in relapse rate. The clinical
rationale of assertive outreach is that im-
proved continuity of care leads to better
clinical management and reduced relapse
(Stein & Test, 1980) and that reduced hospi-
talisation is a reflection of this (Marshall
& Lockwood, 1998). In studies of crisis
resolution/home treatment teams, however,
the intervention comprises a different style
of managing relapses, not preventing them
(Johnson et al, 2005; Glover et al, 2006;
Killaspy et al, 2006). Thus a reduction in
hospitalisation is a marker for more effec-
tive management of relapse (i.e. successful
care in the home) not a marker for reduced
relapse. The relationship between hospital-
isation and relapse in these two different
types of studies needs to be recognised for
their interpretation.

MISINTERPRETATION OF
HOSPITALISATION AS AN
OUTCOME
Hospitalisation as an outcome measure in
community studies draws its legitimacy in
RCTs from the highly plausible assumption
that the threshold applied in any local area
will affect experimental and control groups
equally. Thus any differences in hospitalisa-
tion rates can be attributed to differences
between the two interventions. The dangers
of extrapolating directly from model
services, with their highly motivated staff,
exclusion criteria and invisible incentives,
have been well documented (Bachrach,
1989; Tyrer et al, 1999) although the lesson
is consistently ignored. Within an individual
trial, however, difference in hospitalisation
is generally a reliable guide to anticipated
impacts. The wider generalisability of hospital-
isation is a highly complex matter and failure
to give it due consideration has led to signif-
icient mistakes in policy and planning.

Can hospitalisation rates be used
in meta-analyses?
Meta-analyses of medical trials consolidate
the outcomes from several small trials into
a single result for that outcome, treating
all the data as if from a single trial. The
benefits of this approach, and the world-
wide Cochrane Collaboration that supports
it, is that conclusions can be established
earlier (thereby introducing life-saving
treatments and also avoiding unnecessary
subsequent trials) and with greater confi-
dence. The delay in introducing clot-busting
drugs after myocardial infarction is often
cited as the most convincing case for meta-
analyses (Antman et al, 1992). The
importance of meta-analyses has been em-
phasised for mental health research because
of the preponderance of small, underpow-
ered studies (Coid, 1994). Within the
Cochrane Collaboration, difference in
hospitalisation rates has been the most in-
fuential outcome in meta-analyses of com-
munity mental health services (Marshall
although others are reported (e.g. loss to
follow-up care, satisfaction with care, cost
of care). Clinical and social functioning
are often too inconsistently collected for in-
fuential findings to be presented.

The meta-analyses of hospitalisation
for ACT teams (Marshall & Lockwood,
1998) and case management (Marshall et
al, 2001) have been consistently cited to
confirm that ACT reduces the need for hospi-
talisation compared with standard care.
As a consequence, ACT has been mandated
in many US and Australian states, Cana-
dian provinces and increasingly across
Europe. In the UK ACT teams are the basis
for the reorganisation of mental health
services required by the NHS Plan
(Department of Health, 2000), with the
establishment of over 170 teams. Close ex-
amination of the forest plots indicates that
there is quite a lot of heterogeneity in the
results. Some caution should therefore be
exercised in applying meta-analytical tech-
niques to hospitalisation outcomes and ef-
forts should be made to understand the
source of the heterogeneity.

Two potential sources of heterogeneity
are immediately clear from a cursory ex-
amination of the forest plots. First, in the
ACT meta-analysis the studies demonstrat-
ing major reductions are all from the USA,
and the only non-American study included
(Muijen et al, 1992) demonstrates minimal
reduction. In the case management analysis
three of the studies are from the UK. This
difference might indicate an impact of dif-
fering healthcare systems on the results of
these two meta-analyses. There is also a
suggestion that later studies indicate less
benefit for ACT, although the difference is
not as pronounced as that for the geogra-
phical differences. The importance of these
observations becomes clear with the failure
of any recent, high-quality European studies
of ACT to replicate the reduction in hospi-
talisation. Indeed several recent European
studies have been sufficiently powered that
their failure to demonstrate reduction in
hospitalisation can be interpreted as confir-
mation that there is no reduction. Hospita-
lisation is therefore not a reliable outcome
in meta-analyses. Variation in hospitalisa-
tion as an outcome, on the other hand,
has proved to be most useful by leading
analyses that produce better understanding
in service evaluations.

Control services are not placebos
Examination of the differences between US
and European (predominantly UK) com-
munity care studies confirmed that the
impression that US studies were more
successful in reducing hospitalisation is in-
deed the case (Burns et al, 2002). This holds
despite evidence that the interventions were

s39

HOSPITALISATION AS AN OUTCOME MEASURE
substantially similar (Fiander et al, 2003). Home-based care in the US (the definition was widened to ensure consistency and to avoid post hoc rationalisation in labelling) did reduce in-patient care by a statistically significant mean of about 10 days a year compared with standard care, whereas in European studies it increased in-patient care by a non-significant average of 3 days a year. However, the conclusion that US experimental services kept patients out of hospital more is not supported. Mean days in hospital were essentially the same for experimental service patients in the US and Europe (19 and 21 days respectively); the differences stem from the differences in hospitalisation for the control services (means of 28 and 17 days respectively).

This exploration of variation in hospitalisation data confirms our earlier call for community psychiatry studies to pay much greater attention to service characterisation and, in particular, characterisation of the control services (Burns & Priebe, 1996). Hospitalisation as an outcome measure certainly has some generalisability, but its limitations need to be considered when it is used as a basis for service planning.

**Distinguishing effective ingredients**

An important consequence of the heterogeneity of hospitalisation as an outcome is that it has stimulated a search for the sources of that heterogeneity and this has helped distinguish effective from more redundant components in complex interventions. In the systematic review of home-based care by Catty et al (2002) we obtained data from the 60 of the 90 researchers to characterise their experimental services at the time of the investigations. The information was collected using 20 operationalised ‘components of care’, which were subjected both to cluster analysis to identify common characteristics of practice and to regression against reduction in hospitalisation to identify whether any were more strongly associated. Figure 1 shows the six regularly occurring components reported. The two found in a regression analysis to be significantly associated with reduction in hospitalisation are home visiting and joint health and social care. This is only a post hoc analysis and the sample was quite restricted. However, what it does do is indicate how hospitalisation as an outcome can be used to explore community mental health services in greater depth.

A subsequent study has demonstrated even more convincingly the utility of hospitalisation as an outcome measure to exploit differences in trials (Burns et al, 2005). Meta-regression analysis allows skewed, non-parametric data to be used in a meta-analysis (Thompson, 2001) and allows for multi-site studies to be analysed as, effectively, several independent studies. Substantially the same set of studies as those used in the home-based care review were used to obtain patient-level data. Hospitalisation data were assessed using an accepted model fidelity scale (McGrew et al, 1994). This approach demonstrated that baseline bed use was the factor most strongly associated with reduction, but of the model fidelity factors it was predominantly the structural rather than staffing characteristics of the intensive approach which accounted for the outcome differences.

**CONCLUSIONS**

Hospitalisation as an outcome in schizophrenia research is likely to retain an important place mainly because of its obvious utility to planners and service providers. It also has a powerful advantage in its face validity to clinicians. Its limitations are obvious – it says little about individual patient outcomes and can convey a sense of being more interested in services (in particular their costs) and the professionals that staff them than in patient welfare. A closer examination demonstrates that reducing unnecessary hospitalisation has paralleled patient and carer wishes. Survey after survey has reported the desire to remain out of hospital as much as possible (Drake & Wallach, 1988). Given the choice, patients almost invariably opt for out-patient and community care or, failing that, day care.

Reducing unnecessary hospitalisation has also, arguably, increased the overall efficiency of mental healthcare. The disproportionate cost of in-patient care per patient contact (which is, after all, where the treatment occurs) reflects the capital costs, hotel costs and 24-hour staffing. In-patient care has declined for most physical disorders as the population increasingly has clean, well-heated accommodation allowing adequate privacy. These extra costs of hospital care are justified when they add to safety or ensure adherence. However, for many patients it is not necessary and there is no clear evidence that treatments are any more effective for being delivered in hospitals than in clinics or patients’ homes. Indeed, the difficulty of ‘transfer of learning’ from hospital to home is one of the underlying reasons for Stein & Test’s emphasis on what they call ‘in vivo’ care in assertive outreach (Stein & Test, 1980).

Reducing hospitalisation is also in line with most current thinking in bioethics, where the emphasis has been on the provision of mental healthcare in the ‘least restrictive’ environment (Lin, 2003). Much of this ethical debate has centred around the care of legally detained patients. However, there is accumulating evidence of informal coercion in mental healthcare (Monahan et al, 2005), suggesting that the distinction between voluntary and involuntary may be better conceptualised as a gradient rather than a dichotomy (Bonnie & Monahan, 2005). Patient and ethical views about legally enforced admission may, in some measure, also apply to most admissions.

The utility and apparent simplicity of hospitalisation as an outcome measure

Fig. 1 Consistent care components of home-based care. From Wright et al (2004). Reprinted with permission.
should not, however, blind us to its limitations. It is a good proxy for relapse in schizophrenia in well-functioning and coordinated services. However, it is a social sciences outcome that is not independent of context and it needs to be interpreted that way. Its reputation has been somewhat tarnished by overextrapolation; there is a need for greater caution in its interpretation to ensure its reputation is rehabilitated.

REFERENCES


Health economic measures in schizophrenia research

PAUL McCrone

Background It is essential in economic evaluations of schizophrenia interventions that all relevant costs are identified and measured appropriately. Also of importance is the way in which cost data are combined with information on outcomes.

Aims To examine the use of health economics in evaluations of interventions for schizophrenia.

Methods A review of the key methods used to estimate costs and to link costs and outcomes was conducted.

Results Costs fall on a number of different agencies and can be short term or long term. Cost-effectiveness analysis and cost–utility analysis are the most appropriate methods for combing cost and outcome data.

Conclusions Schizophrenia poses a number of challenges for economic evaluation.

Declaration of interest PM has received speaker fees from Eli Lilly and Janssen-Cilag.

In evaluating interventions to treat schizophrenia it is clearly important to assess the clinical and social effects of the intervention and to measure these appropriately. It will often be desirable to include a number of outcome measures, ranging from those with a relatively narrow perspective (e.g. symptom measures) to those which are more holistic (e.g. quality of life measures). Ideally, we would hope that decision makers would only favour interventions which have proven efficacy. Although good evidence is certainly required for new treatments or services, some of those in existence probably do not have as firm an evidence base.

Evidence that an intervention is effective does not in itself mean that it should be implemented. All interventions require the use of resources (money, time, equipment, etc.), and these will inevitably be limited in their supply and able to be used to provide care in other ways (whether for people with schizophrenia or with asthma, cancer, etc.). Resource scarcity and competing demands for these resources means that the costs of interventions for schizophrenia need to be assessed in addition to outcomes.

MEASURING COSTS

Costs occur when a productive activity takes place that necessitates the use of scarce resources that could be used for some other purpose. Economists describe this as an opportunity cost. The time that psychiatrists, community mental health nurses, psychologists, social workers and others spend in providing care for people with schizophrenia could be spent with other patient groups, and hence a cost arises. Unpaid care (from families/friends) will also frequently be used, and this also has an opportunity cost.

The focus of economic analyses is often on the final figure (in pounds, dollars, euros, etc.), but this is simply a proxy measure for care inputs (and in schizophrenia these will be many) that a patient receives. It is generally incorrect to regard costs as outcomes; rather they are a representation of inputs that produce outcomes. Some of the impacts of interventions for schizophrenia can be measured in monetary units, for example in-patient admissions, if a reduction is the main aim of the intervention, as with crisis resolution teams.

To estimate the cost of schizophrenia care it is necessary to identify all the resources required to provide this care and also those resources that are subsequently affected. The costs of treatment with atypical antipsychotics, for example, include not only the specific drug costs but also the professional time required to administer and monitor treatment plus the cost of other care (such as in-patient stays and outpatient appointments) if these may change as a result of the drug treatment.

A common costing perspective in economic evaluations is to include all health service costs. This is the preferred view of the UK’s National Institute for Health and Clinical Excellence (NICE), whose remit is to offer guidance on the spending of National Health Service finances. In schizophrenia care even this will be inadequate, as interventions which are successful (e.g. in reducing positive symptoms) could alter the use a patient makes of social care, education and criminal justice services. Costs to families and friends in terms of the unpaid care (but representing costly lost opportunities) provided and also the potential value of patient time spent in contact with services could also be included.

The Client Service Receipt Inventory (CSRI; Beecham & Knapp, 2001) is the most commonly used questionnaire to measure service costs for people with schizophrenia. The CSRI is usually completed through interviews with patients, although case notes and information from carers can be used, and some versions can be completed by the patients themselves. The information collected with the CSRI, or a similar measure, can be combined with appropriate unit cost information (e.g. Curtis & Netten, 2006) to generate service costs.

Costing systems of care

Several studies have assessed the economic impact of alternative approaches to delivering schizophrenia care. In a review of home-based care (assertive community
treatment and crisis interventions) Burns et al (2001) identified 22 studies that had included an economic component. It was encouraging that many of these did take a relatively comprehensive approach to costing and that this had improved compared with an earlier review (McCrone & Weich, 1996).

Broader studies have attempted to cost the impact of schizophrenia at a national and even international level. Making comparisons between countries presents difficulties given variations in the methodologies used (Andlin-Sobocki & Rossler, 2003). Differences in price levels generally (and healthcare prices in particular), the supply of specific services and the structure of mental health services will all affect service costs, as will variations in patient need. Nevertheless, it can still be informative to examine the contribution that specific services make to overall service costs.

In-patient care has consistently been the most costly service since deinstitutionalisation began. Knapp et al (2004) reported that it accounts for between 28 and 94% of direct healthcare costs in schizophrenia, whereas medication usually accounts for less than 15% of costs. One exception was in Nigeria where medication accounted for 62% of costs, reflecting the much higher drug costs relative to cost of in-patient care in low- and middle-income countries. This is a vivid reminder that cost findings in one area cannot be simply translated to another. The supply of services will naturally have a major impact on utilisation and therefore costs. In a five-country comparison of schizophrenia care (Table 1) costs were particularly low in a Spanish city where there was limited availability of day or residential care (Knapp et al, 2002).

### Costs of specific interventions

Economic evaluations have been conducted of a wide range of interventions for schizophrenia (McCrone & Weich, 1996; Byford et al, 2003). Most tend to concentrate on healthcare costs, with some also including social care and criminal justice service costs. Very few evaluations have assessed the impact that interventions have on informal care costs, an important omission as many community-based interventions could well increase the necessity for care from family members or friends.

Another neglected cost is patient time spent using services. In economics generally, time is valued using information on an individual’s wages, but the majority of patients in contact with specialist services are not in work. However, this does not mean that the value of time is zero – time spent on other activities will still have to be forgone in order to use services. This might be relevant when evaluating assertive community treatment, which is principally targeted at patients who are ‘difficult to engage’. Although it is probable that engagement problems are usually a result of the clinical features of schizophrenia, some patients might be engaged in other activities that they value and which limit the time which they wish to spend accessing services (even if this is to the detriment of their mental health).

Cost information is also relatively limited concerning the period before patients receive care for schizophrenia. The duration of untreated psychosis (DUP) has been reported to be up to 1–2 years (McGlashan, 1999). Case-note narratives and studies of pathways into care suggest that this period is characterised by contacts with criminal justice services, visits to accident and emergency departments and interruptions to employment. All of these consequences incur costs, and interventions to reduce the DUP should reduce these, but to date no study has calculated this effect. Further discussion of the DUP is provided by Singh (2007) in this supplement.

Most economic (and clinical) evaluations have relatively short time scales. The consequence of this is that the longer-term cost impact of improved, or worsened, outcomes is neglected. If an intervention is successful then it is highly likely that this will have a long-term impact on the use and cost of services, especially for people with a long-term condition such as schizophrenia where some of the care inputs required have particularly high costs. However, the extent of cost savings as a result of successful interventions is unclear. Engagement with services might be enhanced, which could actually increase long-term costs. However, it is probably more likely that effective care will reduce the need for crisis services – and in particular emergency admissions – if effectiveness means fewer and less severe relapses.

### LINKING COSTS TO OUTCOMES

The measurement of costs associated with interventions for schizophrenia forms only one component of an economic evaluation. The synthesis of information on costs with data on effectiveness is the essence of economic evaluation. This is illustrated in Fig. 1 where costs (should) depend on patient needs and influence outcomes. There are different forms of economic evaluation and these differ according to how outcomes are measured.

#### Cost-minimisation analysis

There are some situations where the outcomes of different services are known a priori, and therefore only costs need to be measured. If outcomes are identical then the least-cost option is the preferred one. However, given the variety of possible interventions for schizophrenia (typical and atypical antipsychotics, psychological therapies, early intervention, etc.) only rarely will outcomes be known with confidence beforehand.

#### Cost-effectiveness analysis

Cost-effectiveness analysis combines cost information with data on a single condition-specific outcome measure, for example symptomatology or social functioning. This

<table>
<thead>
<tr>
<th>Service</th>
<th>Amsterdam</th>
<th>Copenhagen</th>
<th>London</th>
<th>Santander</th>
<th>Verona</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital in-patient</td>
<td>320</td>
<td>5772</td>
<td>3659</td>
<td>1456</td>
<td>2705</td>
</tr>
<tr>
<td>Hospital out-patient</td>
<td>236</td>
<td>376</td>
<td>139</td>
<td>0</td>
<td>627</td>
</tr>
<tr>
<td>Day care</td>
<td>2293</td>
<td>774</td>
<td>1091</td>
<td>–</td>
<td>1650</td>
</tr>
<tr>
<td>Community services</td>
<td>551</td>
<td>241</td>
<td>1086</td>
<td>90</td>
<td>459</td>
</tr>
<tr>
<td>Residential care</td>
<td>764</td>
<td>130</td>
<td>749</td>
<td>–</td>
<td>298</td>
</tr>
<tr>
<td>All care</td>
<td>4112</td>
<td>7460</td>
<td>6771</td>
<td>1444</td>
<td>5760</td>
</tr>
</tbody>
</table>

form of analysis can have particular relevance for clinicians. However, it is limited given that schizophrenia affects patients in many ways. This is a common limitation of most clinical evaluations, which generally specify a primary outcome measure.

**Cost–consequences analysis**

This is a more general case of cost-effectiveness analysis that reflects the broad impact that interventions have. Rather than one single outcome measure being used, a number of different measures are considered. Economic evaluations of interventions could combine cost-effectiveness and cost-consequences analyses, with the former conducted using the primary clinical outcome measure and secondary measures simply reported alongside the cost findings.

**Cost–utility analysis**

This is similar to cost-effectiveness analysis but uses a generic measure of outcome. This enables interventions in diverse areas (e.g. cancer and stroke) to be compared in terms of cost–utility with those for schizophrenia. The most common outcome measure in this form of analysis is the quality-adjusted life-year (QALY). Having a generic measure of outcome is of particular importance for those making recommendations across diverse clinical areas. In the UK, NICE bases many of its findings on the cost per QALY achieved by different interventions.

Quality-adjusted life-years are a composite measure of the amount of time spent in a particular healthcare state and the quality of life experienced during that time. The latter is measured on a scale from 0 (death) to 1 (full health). The most common quality of life measure used in economic evaluations is the EuroQol–5D (EQ–5D; Williams, 1995). This consists of five domains (mobility, self-care, usual activities, pain/discomfort and depression/anxiety) and respondents state whether they have no problem, some problems or major problems for each domain. This results in distinct health states to which ‘utility’ or quality of life scores (on a 0–1 scale) are attached. Such scores have been derived from a general population survey where people were asked to compare health states with full health (Dolan et al., 1995).

To date, cost-utility analysis has been used infrequently in studies of care for schizophrenia. Most QALY measures (such as the EQ–5D) focus largely on physical health problems and there are questions about the sensitivity of QALY measures to changes in mental health states (Chisholm et al., 1997). However, there have been a small number of attempts to use the QALY method in studies of schizophrenia care (e.g. Rosenheck et al., 1998; Sevy et al., 2001) and more work in this area would be useful.

**Cost–benefit analysis**

Cost–benefit analysis compares the costs of a particular service with the outcomes achieved also measured in monetary terms. If the outcomes in monetary terms exceed the costs then the service is efficient. One of the earliest economic evaluations of community mental health services in Madison (Wisconsin, USA) used this method, with earnings from work used as the main outcome measure (Weisbrod et al., 1980). The breadth of ways in which interventions for schizophrenia might affect patients suggests that cost–benefit analysis will rarely be appropriate.

**ANALYSIS AND INTERPRETATION OF ECONOMIC RESULTS**

If a new intervention for schizophrenia is compared with usual care then a number of results could occur. It would be appropriate to adopt an intervention if it results in lower costs than existing care and better outcomes. The intervention should also be favoured if outcomes are no different but costs are reduced or if costs are the same and outcomes are improved. Usual care would be the preferred option if the results were the opposite way round. However, it is unclear whether or not a new intervention should be adopted if outcomes are better but costs are higher. In effect this becomes a value judgement, with the key question being whether or not the increased costs are justified by the level of improved outcomes. Similarly, there is ambiguity about the appropriateness of an intervention if it saves money but is less effective. Interestingly, there appear to be few studies in this category. Interventions with significantly better outcomes than comparators will frequently reduce costs (especially those associated with admissions).

Many economic evaluations actually have uncertain results. For example, both the UK700 study (Byford et al., 2000) and the PRISM Psychosis Study (McCrone et al., 1998), which compared the costs and outcomes of intensive and standard care services for people with psychosis, found no statistically significant differences between costs and outcomes.

To address such uncertainty, it is becoming common practice to link cost and outcome data from cost-effectiveness and cost-utility analyses using incremental cost-effectiveness ratios (ICERs). An ICER is defined as the ratio between the difference in costs between two interventions and the difference in outcomes. The ICER then allows us to state the cost that is incurred for an intervention to produce one unit of outcome more than a comparator. An ICER is only informative if one intervention is both more effective and more costly than its comparator (otherwise the intervention or the comparator would be the dominant option). Relating incremental costs to incremental outcomes suggests that a unit improvement in outcome (e.g. a 1-point decrease in score on the Brief Psychiatric Rating Scale or the Positive and Negative Syndrome Scale) has a meaning to decision makers. The latter may not hold and so more tangible outcome measures might be preferred (e.g. measuring outcome in terms of number of clinically important changes or number of people above or below a particular clinical cut-off point).

One of the problems with ICERs is that they rely on point estimates of costs and outcomes and yet there will be uncertainty around both of these. This has led economists to produce cost-effectiveness acceptability curves. These show the probability that one intervention is more cost-effective than another for alternative values (thresholds) placed on a unit improvement in outcome.

**CONCLUSIONS**

Since the 1980s there has been a substantial increase in the number of studies that have examined the costs associated with
schizophrenia and economic evaluations of specific interventions for this disorder. The quality of studies has generally improved but there are still limitations. In particular, recent methodological developments in health economics generally need to be applied to multi-site studies to generate evidence from the large samples that these studies can provide. Schizophrenia affects the lives of patients in many ways and it is essential that outcome measures used in economic evaluations reflect this. What is clear is that the costs of care should be measured comprehensively and over an adequate period. Costs also need to be linked to outcomes appropriately. The most suitable methods for doing this appear to be cost-effectiveness and cost-utility analyses.

REFERENCES


Cognition as an outcome measure in schizophrenia

MICHAEL S. KRAUS and RICHARD S. E. KEEFE

Background  Cognitive deficits are a core feature of schizophrenia. These deficits are not caused by medication or symptoms, and have a dramatic negative effect on real-world functioning.

Aims  To critically examine a selection of the most common batteries used to assess cognition in schizophrenia.

Method  Literature review of cognitive assessment batteries for use in schizophrenia.

Results  A wide variety of neurocognitive test batteries have been developed or adapted to assess cognition in schizophrenia. These differ in time requirements, repeatability, ease of administration, degree of face validity, availability of co-normative data and degree to which results can be parsed into separate domains of cognitive functioning. The most appropriate depends on the setting and the question being addressed.

Conclusions  Cognitive outcome measures have reshaped our understanding of schizophrenia and will be essential tools for unravelling the aetiology of the disease and designing more effective interventions.

Declaration of interest  R.S.E.K. receives royalties from sales of the Brief Assessment of Cognition in Schizophrenia (BACS) battery and the MATRICS Consensus Battery (MCCB). He is a member of the MATRICS Neurocognition Committee and Director of the TURNS Chief Neuropsychologists Group. He receives consultancy fees from several pharmaceutical companies.

In an attempt to classify the multitude of mental disorders he encountered in his work, Emil Kraepelin adopted the term ‘dementia praecox’ to label a condition characterised by early psychosis and cognitive deterioration (Hoenig, 1983). Although Bleuler renamed the disease schizophrenia in 1911, emphasising his view of the disease as a lack of connection between a person’s affect, thought and perception, he still viewed cognitive deficits as integral to the disorder (Gabriele, 2000). The perceived importance of cognition in schizophrenia has since waxed and waned. Cognitive and negative symptoms, which were regarded as integral by both Kraepelin and Bleuler, were later overshadowed by the more easily observable and identifiable positive symptoms. The Research Diagnostic Criteria (RDC), which were designed to formalise the diagnosis of mental disorders, emphasised the Schneiderian symptoms, and this tradition has continued into DSM–III and IV (Andreasen, 1997). Although negative symptoms were added as criteria in DSM–IV, cognition is still not included in the formal criteria.

Although the move away from cognitive impairment as a focus in schizophrenia was initially motivated by enhancement of diagnostic reliability, it came to shape how the disease was viewed and investigated. However, a renewed interest in cognition has been evident recently, spurred in part by the strong empirical relationship between cognition and real-world functioning (Green, 1996). Several studies have failed to demonstrate a significant correlation between positive symptoms and functional outcome (Green, 1996), suggesting that a diagnostic and treatment focus on Schneiderian first-rank symptoms has sidelined key aspects of the disease.

IMPAIRED COGNITION AS A CORE FEATURE

A mounting body of evidence indicates that diminished cognitive ability is a core feature of schizophrenia. Severe impaired performance on cognitive tests (two standard deviations below the mean of healthy controls) in several cognitive domains is strong evidence for the importance of cognitive impairment in the disease (Saykin et al., 1991; Harvey & Keefe, 1997). Broad cognitive deficits, of moderate to severe magnitude, have been found in meta-analysis (Heinrichs & Zakzanis, 1998), large clinical trials (Harvey et al., 2003, 2004; Keefe et al., 2006a) and research studies (Bilder et al., 2000; Heaton et al., 2001; Keefe et al., 2004). Cognitive deficits have been shown to lack correlation with severity of positive symptoms and to be only mildly correlated with severity of negative symptoms (Addington et al., 1991; Gold et al., 1999; Keefe et al., 2006a), indicating that impaired cognition is not an epiphenomenon of clinical symptoms.

Although some studies have indicated that a significant portion of people with schizophrenia test in the normal cognitive range (Palmer et al., 1997), strong evidence suggests that even these exhibit cognitive abilities below those expected if they did not have the disease. A study of monozygotic twins found that 80–95% of twins with schizophrenia scored below their unaffected twin (Goldberg et al., 1993). Another study found that 98% of people with schizophrenia performed below the level predicted by estimates of their premorbid functioning based on level of parental education, compared with 42% of controls (Keefe et al., 2005).

Many early studies of the cognitive deficit in schizophrenia were of people who were either taking antipsychotics at the time of the study or had taken them in the past. However, several studies have since demonstrated cognitive deficits in people with first-episode schizophrenia who have never taken antipsychotics (Saykin et al., 1994; Mohamed et al., 1999; Bilder et al., 2000; Torrey, 2002).

Unlike Schneiderian first-rank symptoms, cognitive deficits correlate highly with measures of functional outcome (Velligan et al., 1997; Addington & Addington, 1999; Green et al., 2000). In addition, the literature overwhelmingly supports a longitudinal correlation between cognitive ability at baseline and later assessments of functional outcome (Green et al., 2004; Carlsson et al., 2006), suggesting that cognitive deficits are a key and perhaps limiting factor in rehabilitation of people with schizophrenia. The overwhelming evidence...
supporting neurocognitive deficits as a core feature of schizophrenia and predictive of functional outcome has spurred the United States National Institute of Mental Health (NIMH) to target such deficits for pharmacological intervention (Marder & Fenton, 2004).

**SEPARATE DOMAINS OF DEFICIT V. GENERAL DEFICIT**

A crucial consideration for practical assessment of cognition in schizophrenia centres around whether the cognitive deficit is best described as broad or is more pronounced in specific domains. A large factor-analytical study involving the Wechsler Adult Intelligence Scale (WAIS–III) and Wechsler Memory Scale (WMS–III) batteries found that the performance of 1250 healthy controls was best described by a model composed of six separable domains of cognition: verbal comprehension, perceptual organisation, auditory memory, visual memory, working memory and processing speed (Tulsky & Price, 2003). Although a meta-analysis of 22 studies of cognitive performance in populations with schizophrenia reported a strikingly broad deficit spanning all domains (Heinrichs & Zakzanis, 1998), some theories have focused on specific domains of impairment, such as working memory (Goldman-Rakic, 1994), verbal memory (Saykin et al., 1994), and executive functions (Goldberg et al., 1987). However, these results may reflect differences in test sensitivity as opposed to true differential ability across domains.

Although some studies have emphasised differential impairment across domains, others indicate that cognitive performance (Keefe et al., 2006a) and cognitive deficits (Dickinson et al., 2004) exhibited by people with schizophrenia are largely mediated through a single common factor, suggesting a generalised cognitive impairment. This ongoing debate has implications for the aetiology of the disease (whether underlying brain abnormalities are local or global) as well as intervention strategies.

**TOOLS FOR MEASURING COGNITION**

A great many tasks have been developed to assess cognition and various batteries comprised of these tasks have been employed in research with populations with schizophrenia (Table 1). The most appropriate battery for a given study will depend upon the questions being addressed, study settings and available resources. A long battery comprised of many different tests has the disadvantage that missing data will be increased (Keefe et al., 2004). In addition, attrition rates and missing data may be higher in those with the most impairment, thus skewing results. Also, in many settings extensive batteries are impractical because of the time requirements placed on staff members administering the tests. However, a longer test battery usually will increase the ability of the data to measure multiple domains of cognition. Thus if the research question involves the efficacy of a treatment intervention for improving cognition, a composite score from a small battery might be sufficient and allow for a larger number of participants to complete the study while requiring fewer staff resources. If, however, the research question involves relative strengths and weaknesses of various cognitive domains, a thorough battery composed of multiple tasks in each domain may be required. If the research question is primarily focused on one cognitive domain, a brief general battery in combination with

---

**Table 1 Advantages and disadvantages of selected cognitive batteries**

<table>
<thead>
<tr>
<th>Battery</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAIS–III and WMS–III</td>
<td>Long history of use allows comparison with previous studies, norms available</td>
<td>Lengthy, not designed specifically for research in schizophrenia</td>
</tr>
<tr>
<td>MCCB</td>
<td>Designed by panel of experts for research in schizophrenia, allows for domain scores with minimal testing, norms available</td>
<td>Many domain scores based on performance on one test</td>
</tr>
<tr>
<td>SCoRS</td>
<td>High face validity, easily administered by clinicians, minimal time requirements, demonstrated correlation with other measures of cognitive and functional outcomes</td>
<td>High subjectivity</td>
</tr>
<tr>
<td>UPSA</td>
<td>High face validity, proxy test of real-world functioning, minimal time requirements</td>
<td>Domain-level analysis not possible</td>
</tr>
<tr>
<td>RBANS</td>
<td>Minimal time requirements, small practice effects, performance correlated with that on WAIS–III and WMS–III.</td>
<td>Lacks measures of important cognitive domains in schizophrenia, significant ceiling effects</td>
</tr>
<tr>
<td>BACS</td>
<td>Minimal time requirements, designed for use in schizophrenia research, high correlation with composite scores from more extensive batteries, minimal practice effects, available in nine languages, norms available</td>
<td>Domain-level analysis is minimal</td>
</tr>
<tr>
<td>BCA</td>
<td>Extreme brevity of administration, high correlation with extensive cognitive battery and functional outcome measures</td>
<td>Domain-level analysis minimal</td>
</tr>
<tr>
<td>Computerised batteries</td>
<td>Automatic administration reduces rater error</td>
<td>Validity of many tests has not been examined</td>
</tr>
<tr>
<td>Psychophysiological tasks</td>
<td>Much of underlying neurobiological circuitry is known, afford quick assessment of schizophrenia-related endophenotypes</td>
<td>Relationship between improvement in these tasks and functional outcome unknown</td>
</tr>
</tbody>
</table>

---

WAIS–III, Wechsler Adult Intelligence Scale – III; WMS–III, Wechsler Memory Scale – III; MCCB, MATRICS Consensus Cognitive Battery; SCoRS, Schizophrenia Cognition Rating Scale; UPSA, UCSD Performance-Based Skills Assessment; RBANS, Repeatable Battery for the Assessment of Neuropsychological Status; BACS, Brief Assessment of Cognition in Schizophrenia; BCA, Brief Cognitive Assessment.
multiple tests of the domain of primary interest might be most appropriate.

**Wechsler Adult Intelligence and Memory Scales**

The Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1997a) and Wechsler Memory Scale (WMS; Wechsler, 1997b) have long been the most widely employed batteries of assessment for IQ and memory in healthy populations. However, the WAIS-III alone requires approximately 100 min for completion in a mixed clinical population (Ryan et al., 1998). For studies of populations with schizophrenia, researchers using these batteries have tended to reduce the number of sub-tests administered to reduce demands on the patients and staff. Byler et al. (2000) used regression analysis to determine the four tests covering all four domains of functioning assessed by the WAIS-III that would best account for the variance in full-scale IQ in a sample of 41 out-patients with schizophrenia. They found that a shortened version of the WAIS-III, consisting of the sub-tests information, block design, arithmetic and digit symbol took only 30 min to administer and accounted for 90% of the variance in the full-scale IQ of the schizophrenia patients. Because of its brevity, the shortened version of the WAIS may have utility as a routine measure of cognition in clinical practice.

**MATRICS Consensus Cognitive Battery**

As part of the NIMH Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS), the MATRICS Consensus Cognitive Battery (MCCB) was constructed to provide a standard battery for the assessment of cognition-enhancing drugs (Nuechterlein & Green, 2006). From more than 90 tests nominated for inclusion, a final battery of ten tests covering seven domains of cognitive functioning was chosen with a view to practicality of administration, high test–retest reliability, small practice effects, small ceiling effects and demonstrated relationship to functional outcome. The MCCB was designed to assess effects of pharmaceutical interventions on cognition in schizophrenia, the battery is suitable for use in cognitive remediation and non-intervention studies of people with schizophrenia. Although the MCCB requires more time to administer than the shortened WAIS-III, it has the potential to provide a more detailed examination of a patient’s cognitive performance.

**Schizophrenia Cognition Rating Scale**

The Schizophrenia Cognition Rating Scale (SCoRS; Keefe et al., 2006b) is an 18-item interview-based assessment which covers all the cognitive domains tested in the MCCB, except social cognition, and takes approximately 12 min to complete. It is administered separately to the patient and to an informant (family member, friend, social worker, etc.) The interviewee is asked to rate the patient’s level of difficulty in performing various cognitive functions on a 4-point scale, with 4 being the most difficulty and 1 being the least. Upon completion of the 18 items, the interviewee is asked to give a global rating of the patient’s cognitive functioning on a scale of 1–10. The interview has been administered to both the patient and the informant, the interviewer ranks the patient on all 18 items, and gives a global score based on the responses of both the patient and informant as well as the interviewer’s observations of the patient.

Initial assessments of SCoRS results have demonstrated high interrater reliability (Keefe et al., 2006b). The administrator’s global rating was shown to be the single SCoRS measure that correlated most significantly with measures of cognition (BACS; Brief Assessment of Cognition in Schizophrenia; Keefe et al., 2004), performance-based assessment of function (UPSA; Patterson et al., 2001) and real-world assessment of function (Independent Living Skills Inventory (ILSI); Menditto et al., 1999). Step-wise regression analysis demonstrated that the interviewee’s global rating accounted for significant variance in real-world functioning as measured by the ILSI beyond that explained by results from the BACS and the UPSA (Keefe et al., 2006b). Because the SCoRS assessment is based on patient and informant reports, it has high face validity.

In addition to its utility as a coping measure with the MCCB in drug trials, the SCoRS is ideally suited for use in the clinic and may thus increase awareness of cognitive deficits in the diagnosis and treatment of people with schizophrenia. Because patient scores have been found to account for little variance in cognitive performance, functional capacity or real-world functioning scores beyond that accounted for by informant ratings (Keefe et al., 2006b), it is possible that informant ratings alone could be collected when an informant has sufficient contact with the patient. The assessment time could then be limited to 15 min. In addition, the SCoRS should be a familiar procedure for clinicians, who should require significantly less training than for batteries involving less familiar cognitive testing procedures. Interrater reliability of the SCoRS should be established before it is used for clinical or research purposes.

**UCSD Performance-Based Skills Assessment**

The University of California San Diego Performance-Based Skills Assessment (UPSA) was developed as a proxy of real-world functioning that is implemented in role-play. The UPSA measures daily living skills by recreating, in a clinical environment, situations a patient is likely to encounter in the real world. The tasks fall into five categories of functional skills: household chores; communication; finance; transportation; and planning recreational activities. The assessment is relatively brief,
Cognition as an Outcome Measure

Repeatability for the Assessment of Neuropsychological Status

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998) is a brief (45 min) assessment originally designed to test cognitive performance in older patients which has shown utility in providing reliable assessment of cognitive performance in populations with schizophrenia (Willk et al., 2002; Weber, 2003). The performance of people with schizophrenia on the RBANS is highly correlated with performance on the much longer WAIS–III and WMS–III batteries (Gold et al., 1999; Hobart et al., 1999). Because it was designed to be administered repeatedly, the RBANS does not suffer from large practice effects. However, because the battery was developed to test for dementia, it is comprised largely of tests of memory, language and visual perception, and may suffer from ceiling effects on some sub-tests when used in people with schizophrenia. The battery also lacks measures of motor, executive and working memory performance, cognitive domains thought to be important in the cognitive impairment observed in schizophrenia. Despite these omissions, the RBANS is an appealing tool for the assessment of cognition in routine clinical practice owing to its relative brevity.

Brief Assessment of Cognition in Schizophrenia

The Brief Assessment of Cognition in Schizophrenia (BACS; Keefe et al., 2004) retains the positive attributes of the RBANS (brevity of administration and scoring, repeatability and portability) and more completely assesses the extent of cognitive impairment over multiple domains thought to be affected by schizophrenia (Table 2). The BACS, available in nine languages, requires approximately 30 min to complete and is devised for easy administration and scoring. The battery is specifically designed to measure treatment-related changes in cognition, and has alternate forms, thus minimising practice effects. The battery includes brief assessments of executive functions, verbal fluency, attention, verbal memory, working memory and motor speed, and generates a composite score that is calculated by summing z-scores derived by comparisons with a normative sample of 400 healthy controls. Its reliability, validity and comparability of forms has been established empirically (Keefe et al., 2004).

The composite score has high test–retest reliability in people with schizophrenia and healthy controls (intraclass correlation coefficients > 0.80). The BACS composite score has been shown to be as sensitive to the cognitive deficits of schizophrenia as a standard 2.5-hour battery (Keefe et al., 2004) and is highly correlated (r=0.84, P < 0.001) with the composite score derived from the CATIE neurocognitive test battery (Keefe et al., 2007). The BACS also has clear functional relevance, as the composite score is strongly related to functional measures such as independent living skills (r=0.45), performance-based assessment of functioning (r=0.56) and interview-based assessments of cognition in people with schizophrenia (r=0.48) (Keefe et al., 2006c). The BACS is well suited to routine clinical administration when a quick assessment of overall cognitive functioning is required.

### Brief Cognitive Assessment

An even shorter battery is the Brief Cognitive Assessment (BCA; Velligan et al., 2004), which was designed to assess cognition in people with schizophrenia in 15 min. Initial assessment of the battery has indicated good test–retest reliability and strong correlation (r=0.72; P < 0.0001) with an extensive 2-hour battery (Velligan et al., 2004). The two batteries showed similar correlations with measures of functional ability. Normative data are available allowing adjustments for practice effects when performing repeated assessments. The extreme brevity of the BCA makes it a strong candidate for routine clinical administration.

### Computerised Batteries

A recent development in cognitive assessment for clinical trials is the availability of computerised test batteries that allow direct data transfer to study databases. These methods minimise rater error and reduce the costs for human quality assurance. However, many of these methods have not been fully validated and therefore results must be evaluated carefully.

### Psychophysiological Tasks

Assessment of psychophysiological tasks is appealing because much of the underlying neurobiological circuitry has been uncovered in animal and human studies. However, care must be taken when inferring an aetiological basis or treatment strategy for schizophrenia from the performance of patients on these tests. The outstanding requirements.
question remains whether interventions that improve the performance of patients on these tasks would have any effect on the whole of cognition or on functional outcomes. Regardless of their utility in uncovering neurobiological underpinnings of the disease, psychophysiological tasks are particularly appealing for use in genetic studies as tools to quickly assess an endophenotype that might reflect a specific genotypic vulnerability to schizophrenia. Some of the most utilised psychophysiological tasks are briefly described below.

**Eye movements**
Several eye movement abnormalities have been associated with schizophrenia. Two of the most prominent are abnormalities in smooth-pursuit eye movements, in which the patient is instructed to maintain foveation of a smoothly moving target, and antisaccade performance, in which the patient is instructed to make a mirror image saccade away from a suddenly appearing visual cue.

**Antisaccade**
Lesion studies in non-human primates have demonstrated the importance of the dorsolateral prefrontal cortex for inhibiting reflexive prosaccades in the antisaccade paradigm (Fukushima et al., 1994). Likewise, converging evidence has suggested that the dorsolateral prefrontal cortex is compromised in people with schizophrenia (Bunney & Bunney, 2000). A review of patients’ performance on the antisaccade task strongly indicates a significant elevation in erroneous prosaccades that is stable over time (Everling & Fischer, 1998). This was recently replicated in a seven site study by the Consortium on the Genetics of Schizophrenia in which the antisaccade performance of 143 people with schizophrenia was compared with that of 195 controls (Radant et al., 2007). All sites found a significant difference in the number of errors (reflexive prosaccades) made by the two groups. In addition, first-degree relatives of people with schizophrenia have demonstrated higher reflexive saccade rates than unrelated controls (Clementz et al., 1994), suggesting that the endophenotype reflects a genetic vulnerability to schizophrenia.

**Smooth-pursuit eye tracking**
Decreased pursuit gain has long been viewed as a characteristic impairment in people with schizophrenia. However, it was recently shown that people with affective disorder displayed an indistinguishable smooth-pursuit gain (Kathmann et al., 2003). Likewise, unaffected relatives of the two groups did not differ in their pursuit gain deficiencies. These results argue against the utility of smooth-pursuit gain as a phenotypic marker reflecting a genotype specific to schizophrenia. However, high rates of catch-up saccades (Sweeney et al., 1994) and anticipatory saccades (Rosenberg et al., 1997) in the smooth-pursuit paradigm appear to be specific to schizophrenia and may offer phenotypic measures for genetic studies.

**Prepulse inhibition and P50**
Prepulse inhibition and P50 are both measures of pre-attentive processing that display impairment in people with schizophrenia and have been thoroughly researched in animal models. However, impairment in both of these paradigms is fairly widespread over various psychiatric populations, thus decreasing the utility of these measures as an endophenotype for schizophrenia (Bart, 2004)

**CONCLUSIONS**
Cognitive impairment is returning to prominence in the conceptualisation and practical assessment of schizophrenia, and is being considered for inclusion in the ICD and DSM diagnostic criteria. A wide variety of batteries currently exist for assessing cognitive impairment in relation to the disease, and cognitive outcome measures are essential as the field moves forward. However, current options are lacking in several regards. Notably, most cognitive outcome measures tend to suffer from low face validity: it is not obvious to patients or caregivers that improvements in performance on these batteries would make a difference to the patient’s quality of life. Therefore, it is essential in many circumstances that an appropriate test of functional outcome is co-administered with a cognitive battery. Interview-based measures and functional proxy measures are based upon measures of intuitive value to patients and clinicians, and are amenable to use in the clinic, where cognitive assessments are far from routine because of the onerous nature of administering most batteries. Also, although modern cognitive batteries strive to include tasks with low practice effects, this confound is substantial. Thus, trials involving repeated assessments of patients require comparison with comparable control groups that have experienced the same testing procedure. Another consideration in the choice of cognitive testing batteries is that differential sensitivity to between-group or over-time differences across individual tasks can cause spurious findings of differential impairment across domains of cognitive functioning. The parsing of cognitive functioning into various domains is weakened by the lack of domain specificity for many cognitive tests (Keefe, 1995). Despite these shortcomings, cognitive outcome measures have reshaped our view of schizophrenia and will be essential for identifying its aetiology as well as designing more effective interventions.

**REFERENCES**


Neuroimaging and other neurobiological indices in schizophrenia: relationship to measurement of functional outcome

JOHN L. WADDINGTON

**Background** As understanding of the pathobiology of schizophrenia increases, the challenge is to relate such measures to outcome at a functional level.

**Aims** To consider our current understanding of how neurobiological variables relate to functional outcome and might constitute outcome measures in their own right.

**Method** Critical appraisal of recent evidence on structural and functional imaging, neurological evaluation, early neurodevelopmental indices, genomics, proteomics, metabolomics and apoptotic mechanisms in relation to outcome.

**Results** Studies conducted prospectively from the first episode of schizophrenia are generating more reliable findings but currently lack predictive power. Prediction of transition from ‘high-risk’ status to first episode has proved somewhat more fruitful, but the gain has been modest and circumscribed.

**Conclusions** Our current level of understanding does not yet allow the generation of predictive models on an individual patient basis. Genomic and metabolomic studies hold particular potential for generating clinically meaningful ‘biomarkers’ but considerable further work is necessary.

**Declaration of Interest** None. Funding detailed in Acknowledgements.

Over recent decades our understanding of the pathobiology of schizophrenia has increased materially. The body of knowledge accumulated gives us some insight into: (a) aetiological factors, in terms of genetic risk variants (Harrison & Weinberger, 2005; Owen et al, 2005; Karayiorgou & Gogos, 2006) and environmental adversities, both biological and psychosocial (van Os et al, 2005; Spaun et al, 2006); (b) putative pathophysiological processes, in terms of developmental disruption in critical neuronal networks (Waddington et al, 1999; Waddington & Morgan, 2001; Harrison & Weinberger, 2005; Stephan et al, 2006) and the role of these early events in ‘kick starting’ a lifetime-trajectory model of the disorder (DeLisi, 1999; Waddington et al, 1999, 2007; Baldwin et al, 2004); and (c) the relationship of such pathophysiology to aspects of psychopathology and cognitive deficit (Flashman & Green, 2004).

However, it is important not to underestimate the superficiality of these insights. Studies are, in the main, cross-sectional in nature, often in people with an established illness, and thus fail to inform directly on the relationship of biological variables to subsequent course of illness at any level. A new generation of prospective studies, particularly those from the first psychotic episode (Keshavan et al, 2005; Waddington, 2005), now allows these relationships to be explored systematically, but have focused primarily on the conundrum of whether the biological variables themselves are static or progressive thereafter and how such biology might relate to the longitudinal characteristics of psychopathology and cognitive deficit.

Thus, a fundamental challenge remains: how do we relate cross-sectional and prospective studies of biological measures not just to psychopathology and cognition but also to functional outcome, even on a population basis. More specifically, in the present context the yet greater challenge is to relate specific neurobiological indices to their individual outcomes. The purpose of this article is to outline current understanding of the relationship between neurobiological variables and functional outcome, both on a population basis and in terms of the individual patient, and whether any neurobiological variables have potential for becoming outcome measures in their own right.

**STRUCTURAL IMAGING**

Magnetic resonance imaging (MRI) has established that subtle but widespread abnormalities of cortical and subcortical brain structure are present in schizophrenia on a population basis (Lawrie & Abukmeil, 1998; Wright et al, 2000; Shenton et al, 2001; Honea et al, 2005). A related technique, diffusion tensor imaging (DTI), is a relatively new modality that can examine, through fractional anisotropy, the microstructure of white matter and, through fibre tractography, aspects of neuronal connectivity in the brain. The majority of a limited number of DTI studies to date (Kanaan et al, 2005) have indicated reductions in white matter integrity which supports a broader interpretation of structural brain pathology in schizophrenia to involve cortical disconnectivity (Stephan et al, 2006).

Cross-sectional relationship to outcome

Studies have sought to relate structural brain pathology on MRI to aspects of outcome on a cross-sectional basis, usually by subdividing patients retrospectively into ‘good’ v. ‘poor’ outcome groups and comparing them on various MRI measures at a single index assessment. For example, while patients of whatever outcome evidenced smaller thalamic volume relative to controls, those with poor outcome (i.e. hospitalised for more than 50% of the total duration of illness and continuously hospitalised over the past 3 years) evidenced enlargement of the lateral and third ventricles and reduced overall grey matter volume, particularly in the prefrontal cortex, whereas those hospitalised for less than 10% of total duration of illness and not hospitalised over the past year did not (Staal et al, 2001). Similarly, patients of whatever outcome evidenced an overall reduction in grey matter volume relative to controls, particularly in the frontal and
temporal lobes, but those with poor outcome, ‘Kraepelinian’ schizophrenia (i.e. continuously hospitalised or completely dependent on others for basic needs, unemployed, with severe negative symptoms and severe thought disorder) evidenced reduced grey matter volume in the temporal and occipital but not in the frontal lobes relative to their good-outcome (i.e. non-‘Kraepelinian’) counterparts (Mitelman et al, 2003).

However, such studies illustrate how the indirect nature and inconsistency of the relationship between MRI measures and retrospective, dichotomous indices of outcome, on a population basis, precludes the generation of a predictive model in relation to the individual patient. Such models can only be generated in longitudinal studies.

**Longitudinal relationship to outcome over chronic illness**

The putative utility of MRI measures as indices of outcome is associated with the enduring debate on the extent to which structural brain pathology in schizophrenia does or does not progress. The weight of evidence from longitudinal studies beginning at various stages of illness (Pantelis et al, 2003), together with incisive, pseudo-longitudinal analyses of cross-sectional studies (Woods et al, 2005), now suggests a small but significant acceleration in the loss of cortical grey matter volume with associated enlargement of cerebrospinal fluid (CSF) spaces in the long term. To the extent that one can generalise from very early-onset schizophrenia in children to the more typical presentation in young adulthood, there is initial evidence for dynamic tissue loss that progresses to frontal, but less so to cingulate and temporal cortex, in ‘waves’ along the anterior–posterior and dorsal–ventral axes (Vidal et al, 2006).

Studies have related such longitudinal changes to aspects of outcome, usually in terms of MRI measures and assessments of clinical course made on two occasions at the beginning and end of a given period of follow-up. For example, over a mean interval of 3.6 years (range 0.6–7.5), patients who had been ill for a mean of 15.3 years (range 2.2–26.5) evidenced greater decline in cortical grey matter and greater increase in both frontal and temporal cortical sulcal and ventricular CSF volumes relative to controls. The rate of expansion in frontal sulci was associated with higher overall positive symptom scores and a longer percentage of time spent in hospital over follow-up; the rate of decrease in grey matter and of sulcal expansion in prefrontal cortex was associated with higher overall negative symptom scores and longer percentage of time spent in hospital; and the rate of decrease in grey matter and of sulcal expansion in temporal cortex was associated with higher overall negative symptom scores (Mathalon et al, 2001).

However, although such studies support the presence of some subtle but poorly understood neuroprogressive process in schizophrenia, they involve people with differing durations of chronic illness at the start of variable periods of longitudinal assessment who are assessed using a limited range of outcome measures, on a population basis. Thus, they are unable to generate a predictive model in relation to individual patients. Such models can only be generated in prospective studies from the first psychotic episode.

**Prospective relationship to outcome from the first episode**

It is now clear that some aspects of structural brain pathology evident on MRI in chronic schizophrenia, particularly reduction in whole brain and hippocampal volumes, together with enlargement of the third and lateral ventricles, are present at the time of the first psychotic episode and therefore pre-date onset of diagnostic symptoms (Steen et al, 2006; Vita et al, 2006). Diffusion tensor imaging also indicates lower fractional anisotropy, consistent with white matter abnormalities at the first episode (Szczesnko et al, 2005). Thereafter, as considered above in the context of chronic illness, there is a small but significant acceleration in loss of cortical grey matter volume with associated enlargement of CSF spaces (Pantelis et al, 2005; Whitford et al, 2006). The ‘anchor’ event of the first psychotic episode provides a frame of reference from which any relationships between MRI parameters and long-term outcome can be explored prospectively.

Studies have related such longitudinal change to prospective evaluation of outcome, usually in terms of MRI measures and assessments of clinical course made on two occasions: at the first episode and at variable periods of follow-up. For example, in a prospective study over 3 years, extent of ventricular enlargement during this period was associated with poor outcome as dichotomised in terms of remission of positive symptoms; extent of reduction in frontal lobe white matter volume and of increase in frontal lobe sulcal CSF volume were associated with greater negative symptom severity; extent of reduction in frontal lobe grey matter volume was associated with poorer executive functioning over follow-up (Ho et al, 2003).

Although this study addressed relationships between longitudinal changes in MRI parameters and outcome measures, a critical question is whether any cross-sectional MRI measure made at the first episode is predictive of subsequent outcome as assessed prospectively. In a prospective study over 2 years, total brain volume and volumes of cortical grey and white matter, third and lateral ventricles and cerebellum at the first or second episode failed to predict outcome in terms of positive or negative symptoms, social disability and need for care (van Haren et al, 2003). Similarly, in a prospective study over 5 years, smaller temporal lobe grey matter volume at the first episode was associated with persistence of hallucinations; however, initial temporal and frontal lobe tissue and sulcal and ventricular CSF volumes were unrelated to negative symptoms, extent of hospitalisation or psychosocial outcome (quality of relationships, sexual activity, recreation and work performance) over follow-up (Milev et al, 2003). In a further prospective study over 2 years, decreasing volume of the dorsolateral prefrontal cortex at the first episode was associated with poorer functional outcome, in terms of social and employment indices, at 1 but not at 2 years (Prasad et al, 2005).

However, although such studies further elaborate the presence of some subtle but poorly understood neuroprogressive process in schizophrenia, they indicate that MRI measures made at and following the first psychotic episode have only limited capacity to predict outcome on a patient population basis. Thus, they are unable to generate a predictive model in relation to individual patients. One potential confound in all such studies is an effect of long-term antipsychotic treatment on brain tissue volumes. Magnetic resonance imaging studies have indicated that volumes of the basal ganglia and pituitary are increased following exposure to typical but not atypical antipsychotics (Lieberman et al, 2005; Pariante et al, 2005). More extensive studies have indicated exposure to typical
antipsychotics to be associated with some reduction in cortical grey matter volume, whereas atypical agents can be associated with some increase in this volume (Dazzan et al, 2005; Garver et al, 2005; Lieberman et al, 2005). Given the practical difficulties in conducting MRI studies during the first episode in a person with no antipsychotic treatment, especially in a clinical setting, these effects could clearly confound the search for relationships between MRI measures and outcome.

**FUNCTIONAL IMAGING**

**Functional magnetic resonance imaging**

Functional magnetic resonance imaging (fMRI) accesses regional neuronal activation in terms of change in blood oxygenation. Although a variety of cross-sectional fMRI findings in schizophrenia have been reported (Tost et al, 2005), systematic, prospective studies from the high-risk state or the first psychotic episode in relation to outcome are in their infancy.

A recent cross-sectional study has examined frontal and cingulate cortex, thalamic and basal ganglia activation during executive processing in those at high risk and in both early-phase and patients with chronic illness in comparison with controls (Morey et al, 2005). The findings indicate that prefrontal function begins to decline before the emergence of diagnostic symptoms and impairment in frontostriatal function is evident thereafter. Recently, a prospective study over a period of 5 years has examined activation during a sentence completion task (Whalley et al, 2006). Cross-sectional fMRI measures made in the high-risk state indicated that those who went on to receive a diagnosis of schizophrenia were distinguished from those who did not by decreased activation of the anterior cingulate, increased activation of the parietal lobe and smaller increases in activation with increasing task difficulty. However, only four high-risk participants evidenced transition to schizophrenia over the follow-up period (Whalley et al, 2006). The elaboration of such fMRI studies in those at high risk and first-episode patients on a prospective basis has the potential to provide important additional information on the prediction of long-term outcome. However, possible confounding effects of antipsychotic drugs on fMRI measures remain a cause for concern (Davis et al, 2005).

**Emission tomography**

Positron emission tomography (PET) and single-photon emission computed tomography (SPECT) are related techniques that access functional processes such as receptor availability, drug-receptor occupancies, transmitter biosynthesis/catabolism and cerebral metabolic activity. Although a variety of cross-sectional PET/SPECT findings in schizophrenia have been reported, only a few studies have systematically applied these techniques prospectively from the high-risk state or the first psychotic episode in relation to outcome. These relate primarily to contemporary versions of the long-standing dopamine hyperactivity model of psychosis, as recently elaborated (Kapur et al, 2005; Seeman et al, 2006).

Over a prospective period of 2 years, higher SPECT binding of $^{123}$I-iodobenzamide to striatal $D_2$ dopamine receptors at the first psychotic episode was associated with poorer social and occupational outcome among those who attained a follow-up diagnosis of schizophrenia but not among those who retained a diagnosis of schizophréniform disorder (Corripio et al, 2006). Recently, a PET study of striatal $^{18}$F-DOPA uptake has indicated dopamine overactivity in people in the high-risk state. These people are being followed prospectively to determine whether those who go
on to evidence a first psychotic episode have raised 18F-DOPA uptake compared with those who do not (Howes et al, 2006). Elaboration of such prospective PET/SPECT studies of those at high risk to include assessment of outcome beyond transition to psychosis has the potential to provide important additional information on prediction of long-term outcome. As for fMRI and MRS, possible confounding effects of antipsychotic drugs on PET/SPECT measures remain a cause for concern (Davis et al, 2005).

NEUROLOGICAL EVALUATION

Electroencephalography

Although electroencephalography (EEG) has a history in biological psychiatry that long pre-dates the emergence of MRI, fMRI, MRS and PET/SPECT techniques, it has failed to match their impact. The recent introduction of techniques such as gamma synchrony has contributed to a new wave of incisive studies.

For example, a cross-sectional study in first-episode schizophrenia has reported decreased magnitude and delayed latency for frontal gamma 1 but not gamma 2 synchrony time-locked to target auditory stimuli, indicating disturbance in connectivity of neural activity in early sensory response to task-relevant stimuli, in a manner that may be modulated by antipsychotic drugs (Symond et al, 2005). The extension of such EEG studies to those at high risk and first-episode patients on a prospective basis has the potential to provide important additional information on prediction of long-term outcome.

Neurological soft signs

Neurological soft signs are non-localising abnormalities that cannot be related to impairment in a specific brain region and are not part of a well-defined neurological syndrome. They constitute evidence for otherwise unspecified brain dysfunction and have been shown consistently to occur to excess in schizophrenia (Bombin et al, 2005).

Studies have indicated that the extent of neurological soft signs at the first episode has little relationship to long-term outcome, for example global or occupational functioning (Bombin et al, 2005). Recently, in a study over a prospective period of 4 years from the first episode, improvement in neurological soft signs score was associated with better overall outcome over the same period (Whitty et al, 2006). However, in a study over a prospective period of 1 year, extent of neurological soft signs at the first episode failed to predict outcome in terms of psychopathology or rate of relapse, defined as hospitalisation or unscheduled visit due to exacerbation, but did predict emergence of tardive dyskinesia (Emsley et al, 2003). Similarly, in a study over a prospective period of 3 years, extent of neurological soft signs at the first episode failed to predict outcome in terms of relapse or occupational functioning (Chen et al, 2005). Thus, the findings to date do not indicate that neurological soft signs exert material prediction of long-term outcome.

EARLY NEURO-DEVELOPMENTAL INDICES

Minor physical anomalies are slight anatomical malformations of body regions that share the ectodermal origins of the brain. Their presence indicates adverse events acting over the first or second trimester (Waddington et al, 1999). Thus, although found reliably to be overrepresented in schizophrenia (McNeil et al, 2000), minor physical anomalies occur to excess in most disorders of early neurodevelopmental origin and therefore likely constitute a non-specific, qualitative indicator of early biological adversity that bears little specific relationship to outcome in schizophrenia.

Anthropometrics of craniofacial dysmorphogenesis

Aspects of dysmorphogenesis, particularly of craniofacial regions that bear the most intimate embryological relationship with early brain development, can be measured using classic anthropometric techniques (Lane et al, 1997). However, in a recent prospective study over a period of 5 years, cross-sectional measurement of hyperтелорism made in the high-risk state did not distinguish those who went on to receive a diagnosis of schizophrenia from those who did not (Johnstone et al, 2005).

Three-dimensional surface imaging of craniofacial dysmorphogenesis

It is now possible to apply three-dimensional surface imaging technology and geometric morphometrics to the quantitative measurement of facial dysmorphology (Hennessy et al, 2005). Such techniques, which have been shown recently to distinguish people with schizophrenia from controls (Hennessy et al, 2007), might have greater potential for addressing any relationship between dysmorphogenesis and long-term outcome in schizophrenia.

Dermatoglyphics

Another index of dysmorphogenesis is dermatoglyphics; for example, a-b ridge count, a quantitative dermatoglyphic measure of the palm, is reduced in schizophrenia (Bramon et al, 2005). However, in a recent prospective study over a period of 5 years, cross-sectional measurement of dermatoglyphics made in the high-risk state did not distinguish those who went on to receive a diagnosis of schizophrenia from those who did not (Johnstone et al, 2005).

Postnatal MRI

Although brain MRI in the immediate postnatal period may have some utility in predicting neurodevelopmental outcome at 2 years (Woodward et al, 2006), any ability to predict schizophrenia as an outcome in young adulthood remains unexplored.

GENOMICS

Evidence accumulated over recent years indicates that schizophrenia is an oligogenic rather than a single-gene disorder, with several risk genes of small effect having been identified (Harrison & Weinberger, 2005; Owen et al, 2005; Karayiorgou & Gogos, 2006).

Studies of association between particular genetic risk variants and aspects of structural brain pathology on MRI in adult schizophrenia are now available (e.g. Cannon et al, 2005; Gurling et al, 2006; Ho et al, 2006). However, only recently are predictive studies emerging. For example, in a prospective study over a period of 5 years, neuregulin 1 and catechol-O-methyltransferase (COMT) genotypic variants assessed in the high-risk state each increased risk for developing psychotic symptoms of schizophrenia in association with abnormalities of brain structure and function on MRI and fMRI (Hall et al, 2006; McIntosh et al, 2006). Such studies may be of considerable heuristic value in relation to prediction of outcome.
Both proteomic and metabolomic studies in schizophrenia are in their infancy, with metabolomics (i.e. the study of the repertoire of biochemcials present in cells, tissue and body fluids as encoded by the genome and modified by environmental factors; Kaddurah-Daouk, 2006) now being explored in the search for biomarkers for several aspects of schizophrenia.

In an initial study using nuclear magnetic resonance (NMR) spectra of CSF samples from antipsychotic-naive or minimally treated patients with first-episode schizophrenia, the glucoregulatory metabolic profile was characteristically altered relative to controls and showed some association with ‘normalisation’ with effective antipsychotic treatment (Holmes et al, 2006). As for genomics, such studies of metabolomics may be of considerable heuristic value in relation to prediction of outcome.

Apoptosis, a form of programmed cell death, is regulated by a complex cascade of pro- and anti-apoptotic proteins that may be altered in schizophrenia and might mediate subtle, progressive loss of cerebral grey matter, particularly over the early course, in the absence of evidence for any neurodegenerative process as currently conceptualised (Waddington et al, 1999, 2007; Glantz et al, 2006). Although a recent study noted apoptotic mechanisms in dermal fibroblasts to be anomalous in schizophrenia (Catts et al, 2006), any ability to predict outcome remains unexplored. However, such indices join genomics and metabolomics in being of considerable heuristic value in relation to prediction of outcome.

CONCLUSIONS

On a population basis, neuroimaging and other neurobiological studies conducted prospectively from the first episode of schizophrenia have advanced rapidly. They hold out the prospect of more reliable findings with further experimental refinement, but currently lack predictive power.

Prediction of transition from ‘high-risk’ status to first episode has proved somewhat more fruitful, perhaps because of contemporary concentration of resources on this critical phase of illness and associated disorders in the context of the potential of early intervention to ameliorate such transition, but the gain has been small. The greatest challenge is to relate specific neuroimaging and other neurobiological indices to outcome on an individual patient basis.

At this stage in our understanding of the biology of schizophrenia over its life-time trajectory, the inconsistency and extent of variability in essentially all such measures still precludes generation of predictive models that are utilitarian for individual patients. Recent fMRI, MRS, genomic and metabolic studies hold the greatest potential for identifying clinically meaningful ‘biomarkers’, but considerable further work is necessary.

ACKNOWLEDGEMENTS

The author’s studies are supported by Science Foundation Ireland, the Health Research Board and the Health Service Executive – North Eastern Area.

REFERENCES


Whitty, P., Clarke, M., McTigue, O., et al. (2006) Diagnostic specificity and predictors of neurological soft signs in schizophrenia, bipolar disorder and other psychoses over the first 4 years of illness. Schizophrenia Research, 84, 110–117.


Outcome measures in early psychosis
Relevance of duration of untreated psychosis

SWARAN P. SINGH

Background  Duration of untreated psychosis (DUP) is considered an important predictor of short-term clinical outcome.

Aims  To explore the evolution of the concept of DUP, synthesise the evidence for its predictive value, highlight the problems in measurement, and consider the potential pitfalls of using DUP as a measure of the effectiveness of early intervention services.

Method  A survey of recent literature was conducted.

Results  Several studies and two systematic reviews confirm that DUP has a robust but moderate effect on outcome in schizophrenia. Studies vary widely in how DUP is defined and measured; hence identifying precise time points when psychosis emerges and remits is conceptually ambiguous and clinically difficult.

Conclusions  Standardised measurement of DUP is a vital first step in allowing comparisons between studies. Duration of untreated psychosis is a relevant measure only of the early detection function of early intervention services.

Declaration of interest  S.S. has received honoraria and educational grants from several pharmaceutical companies.

In the past two decades, the duration of untreated psychosis (DUP) has been an intense focus of clinical and research interest, with the recognition that not only is long DUP associated with poor outcome, but that as a potentially malleable prognostic factor, reducing it at a population level might have a significant public health impact. The suggestion that untreated psychosis may itself be toxic, contributing to a neurodegenerative process, also opens up the possibility of better understanding of the pathophysiology of schizophrenia and the impact of treatment in ameliorating such pathology (McGlashan, 2006). Several studies and two systematic reviews have demonstrated that DUP is an independent predictor of the likelihood and magnitude of recovery in first-episode schizophrenia, although with a small-to-moderate effect (Marshall et al, 2005; Perkins et al., 2005). This paper will explore the historical evolution of the concept of DUP, summarise the evidence on whether DUP is a marker or a determinant of the course of schizophrenia, identify the potential clinical and research problems in defining and measuring DUP, and consider whether DUP is appropriate as a measure of the effectiveness of early intervention services.

By the mid 1980s, studies had begun to demonstrate the importance of the time period between the onset of psychosis and initiation of treatment in determining outcome in schizophrenia (Johnstone et al, 1986; Rabiner et al., 1986). The Northwick Park Study of first-episode schizophrenia found that the most important determinant of relapse was duration of illness prior to starting antipsychotics (Johnstone et al., 1986). ‘Length of manifest illness’ came to be seen as an important contributor to the heterogeneity of outcome in schizophrenia (McGlashan, 1988). It was Wyatt’s seminal review of antipsychotics and the natural course of schizophrenia that firmly established the importance of the length of untreated psychosis as a prognostic indicator (Wyatt, 1991). Wyatt also speculated that untreated psychosis might itself be biologically toxic. Length of untreated psychosis/manifest illness merged into the construct of DUP, and the importance of measuring it developed alongside the first-episode services and research programmes burgeoning across several continents.

DOES DUP DETERMINE OUTCOME?

If ‘untreated’ in DUP refers to pharmacological treatment then, before the introduction of antipsychotics in the 1960s, DUP was the length of the psychotic illness itself. Moreover, if progressive deterioration in untreated patients was inevitable, there would have been an evident improvement in the outcome of psychosis after the introduction of antipsychotics. Recovery rates did increase in the mid-20th century but antipsychotics have not made a huge difference to the proportion of patients in remission in long-term follow-up studies (Hegarty et al, 1994). So does DUP matter?

Besides the humane reasons for reducing DUP and ameliorating unnecessary suffering, there does appear to be a robust, if moderate, effect of long DUP on poor outcome in schizophrenia. In a systematic review of 26 first-episode studies, Marshall et al (2005) found that although a longer DUP was not associated with worse symptoms or poorer functioning at first presentation, at 6 and 12 months following treatment longer DUP was associated with more severe overall symptoms and with worse overall functioning. People with longer DUP were also less likely to experience remission at 6, 12 and 24 months. In a similar review of 43 publications, but using a different meta-analytical strategy, Perkins et al (2005) found that at first presentation longer DUP was associated with more severe negative but not positive symptoms or neurocognitive functioning, and with lower levels of symptomatic and functional recovery from the first episode. Shorter DUP was thus associated not only with greater ‘treatment responsiveness’ but also with greater reduction in negative symptoms, an interesting finding given that negative symptoms are considered less responsive to antipsychotics than positive symptoms. A recent study has even attempted to enumerate the precise effect of DUP on outcome, reporting that each unit increase in DUP is associated with a
The relationship between long DUP and poor outcome is not linear: very long DUP does not correlate with extremely poor outcome. Deterioration in schizophrenia is also unlike that in Huntington’s chorea or Alzheimer’s disease in that it does not go ‘all the way’ but rather reaches a plateau after a few years (McGlashan, 2006). A study of untreated patients from Chennai, South India (Tirupati et al, 2004) found that treatment response is evident even in patients with a DUP of longer than 15 years. In this study, a DUP of less than 5 years predicted good clinical but not occupational outcome, although occupational outcome in such regions is influenced by the family and societal response to the illness rather than being a core feature of the disability itself. So can DUP be meaningfully dichotomised to predict good v. poor outcome?

So far, no demonstrable relationship has been confirmed between effect size of DUP on outcome and the cut-off point chosen to define long or short DUP. Different studies have identified different cut-offs. One study suggested that intensive treatment enhances outcome only if the DUP is less than 6 months (Carbone et al, 1999) whereas another reported that outcomes were significantly worse when DUP exceeded 3 months (Harrigan et al, 2003). Functional outcome appears to decline substantially even after very short treatment delays (>7 days), with more gradual deterioration in functioning up to a very long DUP (>1 year; Harrigan et al, 2003). There does not appear to be a cut-off point associated with medium- to long-term impairment, with some domains of outcome more sensitive to treatment delay than others.

Marshall et al (2005) have suggested that the cut-off point must be very close to onset of psychosis to demonstrate a dichotomous effect. However, the application of a very short cut-off leads to confounding between DUP, outcome and diagnosis (such as acute and transient psychotic disorders). McGlashan (2006) has postulated ‘a window of deterioration’ in the late prodromal phase when neurocognitive decline in particular occurs. Findings that brain abnormalities pre-date frank expression of psychosis make a strong case for intervention in the prodrome of psychosis (Zipursky et al, 1998; Pantelis et al, 2003). However, we are no clearer as to the ‘critical period’ of DUP, exceeding which inevitably predicts poor outcome.

IS DUP CONFOUNDED?

Does the treatment of psychosis treat symptoms alone or the underlying neuropsychological processes? In the latter case, the association between DUP and outcome is easy to understand. In the alternative analysis, shorter DUP is a reflection of prognosis. In this concept DUP is confounded by personality and/or illness-related variables, with a combination of insidious onset, negative symptoms and premorbid dysfunction contributing to delayed help-seeking, delayed initiation of treatment and poorer outcome. Long DUP and later treatment are thus a consequence rather than a cause of other indicators of poor prognosis (Barnes et al, 2000; Verdoux et al, 2001). These confounding prognostic indicators include among others age at onset, gender, premorbid functioning, socio-economic status and mode of onset (Norman et al, 2001; Perkins et al, 2005). Moller (2000) found that a later prodrome onset (mean age in their sample 20.5 years), a prodrome shorter than 2 years, acute initial development of psychosis, the initial presence of grandiosity and/or disorganisation, and a mild level of withdrawal all reduce treatment delay. Mode of onset and premorbid functioning therefore represent built-in components of psychotic illnesses related to a shortened DUP, irrespective of efforts at early intervention.

There is evidence that, even after adjusting for the effects of such confounders, DUP is a significant predictor of outcome. Acute onset, although associated with a shorter duration of initial episode, is not an independent predictor of outcome in psychosis when gender and premorbid functioning are controlled (Singh et al, 2004). Loebel et al (1992) reported that whereas good premorbid functioning is related to higher levels of remission but not a shorter time to remission, shorter DUP correlates with both, suggesting that DUP is a stronger prognostic indicator, perhaps being independent of premorbid functioning. Several other studies that have controlled for premorbid functioning or mode of onset have consistently found that DUP is an independent predictor of outcome (Verdoux et al, 2001; Harrigan et al, 2003; Addington et al, 2004; Melle et al, 2004). Marshall et al (2005) concluded that, although DUP and outcome may be confounded by some third variable, at least in their meta-analysis, premorbid adjustment was not that third variable.
for young people considered ‘at risk’ of impending psychosis. In a series of papers, they have described and refined the concept of ‘at-risk states’ (Yung & McGorry, 1996; Yung et al, 1998; Phillips et al, 2000). The at-risk states include a combination of familial risk (positive family history of psychosis), recent-onset drop in functioning, attenuated or sub-threshold symptoms and brief limited intermittent psychotic symptoms (BLIPS). Attenuated symptoms differ from frank psychotic symptoms in their intensity, frequency and/or duration. Brief limited intermittent psychotic symptoms are frank psychotic symptoms (delusions, hallucinations or thought disorder) which are unequivocally present but last for less than 1 week, resolving spontaneously. Between 30 and 40% of people presenting with such at-risk states make a transition to psychosis, usually within 6 months. A structured instrument is now available, the Comprehensive Assessment of At-Risk Mental States (CAARMS; Yung et al, 2005), which identifies a group at ultra-high risk of making a transition to psychosis. The use of such measures in the general population is problematic since the positive predictive value becomes extremely low in that setting (Warner, 2002). In addition the positive prediction of psychosis appears to be falling in the new prodromal services that are being established (Riecher-Roessler, 2006). This may partly be because such services detect at-risk individuals even earlier (longer duration of prodrome), when emergence of psychosis is more distant in the future and symptoms are even more non-specific.

The PACE clinic is a well-developed and established service that can provide care for large enough numbers of people to allow prospective studies in an at-risk group. However, in most early intervention programmes patients enter the study/treatment after the development of psychosis and DUP can only be measured retrospectively. Several conceptual and methodological problems hinder the measurement of DUP retrospectively. Should onset of psychosis be the onset of any BLIP or the onset of psychotic symptoms that last more than 1 week? Is it possible to make a retrospective judgement of when a sub-threshold symptom crosses the threshold? Can people precisely recall the severity and duration of symptoms that first appeared some months/years ago? What about individuals who report quasi-psychotic symptoms even in childhood, with no identifiable time point where schizotypal traits make a transition to a psychotic state (Poulton et al, 2000)?

**When does DUP end?**

The end of the period of untreated psychosis is conceptually simpler to date, but ‘the start of treatment’ is in reality a similarly complex construct. Does ‘untreated psychosis’ end when any treatment begins, when antipsychotics are started, when antipsychotic treatment at an adequate dose has been adhered to for an adequate period, or when psychosis itself remits? Many studies do not make these distinctions clear in their measure of DUP and scales often do not include a precise definition of treatment adequacy. In routine practice, clinicians sometimes initiate antipsychotics in the prodromal stages of psychosis (Singh et al, 2005a). How should DUP be measured in such cases? In psychosis with prominent mood symptoms at onset, should treatment with antidepressants or mood stabilisers without antipsychotics be considered treatment and hence the end of the period of untreated psychosis? However, what about non-pharmacological treatments which may or may not have an impact on the transition into psychosis (Morrison et al, 2006)?

**Structured DUP assessment scales**

Some scales have been developed to retrospectively map the onset of psychosis. Beiser et al (1993) derived a checklist of behaviours describing the evolution of first noticeable symptoms, emergence of psychosis and initiation of treatment-seeking. The Interview for the Retrospective Assessment of the Onset of Schizophrenia (IRAOS; Hafner et al, 1992) is a semi-structured interview to assess symptoms, psychological impairment and socio-demographic characteristics in the time course of emerging psychosis. The Nottingham Onset Schedule (Singh et al, 2005a) is a short, guided interview and rating schedule to measure onset in psychosis. Onset in the Nottingham Onset Schedule is defined as the time between the first reported/observed change in mental state/behaviour and the development of psychotic symptoms. Onset is conceptualised as comprising: (a) a prodrome of two parts (a period of ‘unease’ followed by ‘non-diagnostic’ symptoms); (b) appearance of psychotic symptoms; and (c) a build-up of diagnostic symptoms leading to a definite diagnosis. The Nottingham Onset Schedule provides a standardised and reliable means of recording early changes in psychosis and identifying relatively precise time points for measuring several durations in emerging psychosis. By varying the starting point of onset, it also allows for several ways of defining and measuring treatment delays, including duration of untreated illness (from start of prodrome to treatment), duration of untreated emergent psychosis (from first psychotic symptom to treatment) and duration of untreated manifest psychosis (from appearance of fully developed psychotic syndrome to treatment).

Table 1 shows a selected number of studies, chosen to demonstrate differences in how the start and end of the period of untreated psychosis are defined and its length ascertained. Given the problems of measuring DUP, it would be surprising if a degree of enforced ‘spurious precision’ did not creep into its measurement. Marshall et al (2005) found that only 12 out of 26 studies included in their review reported a systematic method to assess DUP, with only 5 using a structured instrument (IRAOS in 4, the Royal Park Multidisciplinary Instrument for Psychosis in 1) (Marshall et al, 2005). Overall mean DUP in their meta-analysis was 124 weeks (or 103 weeks excluding an outlier with mean DUP of 796 weeks).

**IS DUP A VALID MEASURE OF EFFECTIVENESS OF EARLY INTERVENTION SERVICES?**

In the UK, the Department of Health has set a national early intervention target of reducing DUP to a service median of 3 months and an individual maximum of 6 months (National Institute for Mental Health in England, 2006). Notwithstanding the methodological problems in measuring DUP, there is the much larger question of whether DUP is an appropriate measure of the effectiveness of early intervention services.

How early is early intervention? It can mean improving outcomes in people with established psychosis by facilitating and consolidating recovery, detecting hidden morbidity in the community by identifying untreated cases of the disorder, or preventing the emergence of psychosis through pre-psychotic interventions. These are different aims, requiring different service models and strategies, and having differing weights of evidence supporting their use (Singh & Fisher, 2000).
Table 1  Selected studies showing variations in definition and measurement of duration of untreated psychosis

<table>
<thead>
<tr>
<th>Study</th>
<th>Start of DUP</th>
<th>End of DUP</th>
<th>Structured assessment</th>
<th>DUP, weeks: mean (median)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addington et al (2004)</td>
<td>First appearance of a positive symptom rated ≥ 4 on the PANSS that lasted throughout the day for several days or several times a week</td>
<td>Onset of first effective treatment</td>
<td>IRAOS</td>
<td>84.2 (28)</td>
</tr>
<tr>
<td>Browne et al (2000)</td>
<td>Time of emergence of psychotic symptoms as dated by patient on basis of SCID interview</td>
<td>Initiation of treatment</td>
<td>No</td>
<td>90.8 (26)</td>
</tr>
<tr>
<td>Carbone et al (1999)</td>
<td>Time of onset of first psychotic symptoms</td>
<td>Entry into treatment programme that includes administration of antipsychotic medication</td>
<td>RPMIP</td>
<td>EPPIC 25.0 (7.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pre-EPPIC 32.4 (4.3)</td>
</tr>
<tr>
<td>Haas et al (1998)</td>
<td>A review of all available sources (interviews with patients, family members, treating clinicians and medical records) were used by two senior clinicians in order to provide a best estimate of time of onset of first psychotic episode</td>
<td>First antipsychotic medication</td>
<td>No</td>
<td>74.4</td>
</tr>
<tr>
<td>Hafner et al (1993)</td>
<td>Onset of first-rank symptoms or meeting criteria for a syndrome based on patient interview</td>
<td>Hospitalisation</td>
<td>IRAOS</td>
<td>109.2</td>
</tr>
<tr>
<td>Ho et al (2000)</td>
<td>Occurrence of delusions, hallucinations, bizarre/ disorganised behaviours, formal thought disorder, or catatonic behaviour at moderate or greater severity</td>
<td>Initiation of antipsychotic treatment</td>
<td>CASH</td>
<td>60.8 (13.5)</td>
</tr>
<tr>
<td>Larsen et al (1996)</td>
<td>Score of ≥ 4 on PANSS positive sub-scale or inappropriate or bizarre behaviour for several weeks</td>
<td>Hospitalisation for psychosis or initiation of antipsychotics for sufficient time and dosage that would lead to clinical response in average patient with non-chronic illness (e.g. haloperidol 5 mg/day for 3 weeks)</td>
<td>No</td>
<td>114.2 (26)</td>
</tr>
<tr>
<td>Malla et al (2002)</td>
<td>Onset of first psychotic symptoms contiguous with presenting episode</td>
<td>Having received antipsychotic therapy for 2 months unless significant response to medication was achieved earlier</td>
<td>No</td>
<td>44.6</td>
</tr>
<tr>
<td>Scully et al (1997)</td>
<td>Age at first admission to a psychiatric hospital</td>
<td>Age at first prescription of antipsychotics</td>
<td>No</td>
<td>722.8</td>
</tr>
<tr>
<td>Singh et al (2005b)</td>
<td>Onset of psychotic symptoms</td>
<td>Commencement of antipsychotics with adherence (at least 75% of prescribed dose taken for at least 75% of the time)</td>
<td>NOS</td>
<td>25.5 (7.4)</td>
</tr>
<tr>
<td>Szymanski et al (1996)</td>
<td>The first time at which psychotic symptoms were noticed by the patient, family or others in the context of a decline in functioning</td>
<td>Entry into research study involving administration of antipsychotics based on clinical judgement of treating physician</td>
<td>No</td>
<td>166.4</td>
</tr>
<tr>
<td>Wiersma et al (1998)</td>
<td>Estimates of psychosis onset were based on WHO structured instruments</td>
<td>Initiation of any form of treatment (almost always involving medication)</td>
<td>Life Chart Schedule and WHO Past and Follow-up History</td>
<td></td>
</tr>
</tbody>
</table>

DUP, duration of untreated psychosis; PANSS, Positive and Negative Syndrome Scale; IRAOS, Interview for the Retrospective Assessment of the Onset of Schizophrenia; SCID, Structured Clinical Interview for DSM–IV; RPMIP, Royal Park Multidiagnostic Instrument for Psychosis; EPPIC, Early Psychosis Prevention and Intervention Centre; CASH, Comprehensive Assessment of Symptoms and History; SOS, Symptom Onset in Schizophrenia; NOS, Nottingham Onset Schedule; WHO, World Health Organization.
1. As reported by Larsen et al (1996).
2. Interquartile range 2–24 months.
3. Long-stay patients with some admitted before the antipsychotics era.
The Lambeth Early Onset (LEO; Craig et al, 2004) and OPUS (Petersen et al, 2005) trials provide convincing evidence that specialised early intervention teams are more effective than standard care in improving clinical outcomes, satisfaction and treatment adherence. Two relatively recent studies provide some justification for prodromal interventions. A randomised controlled trial in a high-risk prodromal population found that a combination of risperidone and psychotherapy reduced the risk of transition to psychosis (McGorry et al, 2002). Pantelis et al (2003) found that specific brain changes accompany prodromal decline and predate the emergence of frank psychosis. If replicated, such studies will provide compelling justification for intervening in the prodromal phase and if all those in the prodromal phase can be identified and adequately treated, there might be no DUP left to measure.

For now at least, the task of reducing DUP at the community level falls on early detection services, which seek undetected cases of established psychosis rather than those at risk. Such programmes usually involve improving knowledge of psychosis within the community and facilitating access to specialist early intervention services. Conducting a randomised trial of early v. late detection is unethical. Hence quasi-experimental designs have been employed to evaluate whether early detection can reduce DUP and improve outcomes while the nature of treatment remains unchanged (Malla et al, 2005). The TIPS project in Norway and Denmark attempted this by comparing outcomes in people with first-episode psychosis who were recruited via an early detection team, with those accessing treatment in an area without early detection but with similar healthcare otherwise. Although the early detection programme did not appear to identify and recruit a large number of previously undetected patients, those entering through this route had shorter DUP (median 5 v. 16 weeks) and better clinical outcomes at 3 months (Melle et al, 2004; Friis et al, 2005; Johannessen et al, 2005). However a ‘before and after’ comparison following the establishment of an early detection programme in London, Ontario, did not find any reduction in DUP in patients recruited following the early detection initiative (median DUP 24.3 v. 21.9 weeks; Malla et al, 2005). Surprisingly patients recruited after the introduction of the early detection programme were more severely ill and had a longer prodromal period.

So far, therefore, the evidence for the effectiveness of early detection programmes in reducing DUP appears limited. Malla et al (2005) have argued that increasing general practitioner knowledge of psychosis and increasing their relative ‘comfort’ in prescribing novel antipsychotics leads to people with milder illness being treated at primary care level, and hence an underestimate of the effect of early detection on DUP. Neither the Scandinavian nor the Canadian study confirmed the presence of a large pool of people with undetected psychosis in the community. This further confirms that DUP is skewed by a small group of outliers with extremely long DUP. Median DUP in well-established community services appears to be relatively low: 32 days for schizophrenia in Nottingham, UK, in one study (Singh et al, 2005a). Although there may be a case for developing targeted early detection programmes for the small proportion of people with undetected psychosis in the community, the evidence for establishing early detection services is not overwhelming.

CONCLUSIONS

Long DUP is clearly associated with poor outcome, independent of the confounders so far explored. Early intervention of specialist services in first-episode psychosis does improve outcomes in the short to medium term. Prodromal services, although potentially very exciting and innovative in creating avenues for treating people who seek help and are at high risk of developing a serious mental illness, are not focused on reducing DUP but on preventing transition to psychosis. Given that we are not able to change prognostic factors such as gender, family history and age at onset, DUP is a malleable variable which should and perhaps can be reduced. However, two caveats remain. First, to make studies comparable a consensus, reliable and replicable measure of DUP should be used across studies to reduce the variation introduced by the measurement process; DUP is a complex enough construct anyway. Second, DUP is not a valid measure for establishing the effectiveness of early intervention services that aim solely to provide evidence-based care in an assertive manner without an early detection arm. Most early intervention services do not conduct early detection, and their effectiveness and rationale should be judged on different criteria: those of meeting a clinical need early, comprehensively and with the best possible available combination of psychosocial and biomedical interventions, rather than simply the reduction of DUP.

REFERENCES


One year outcome in first episode psychosis: impact on duration of untreated psychosis.


Adverse effects of antipsychotics as outcome measures

SAMANTHA HAMER and PETER M. HADDAD

Background Antipsychotic drugs are associated with adverse effects that can lead to poor medication adherence, stigma, distress and impaired quality of life.

Aims To review the use of adverse effects of antipsychotic drugs as outcome measures, with a particular emphasis on methodological issues.

Method Review of data on adverse effects from sources including randomised controlled trials (RCTs), post-marketing surveillance and naturalistic studies.

Results All have advantages and disadvantages and the best overview comes from considering all sources of data together. Adverse effects are inconsistently reported, hampering cross-study comparisons. Many outcome measures lack clinical meaning. In both naturalistic studies and RCTs adverse effects often account for less treatment discontinuation than lack of efficacy.

Conclusions Standardisation in the reporting of adverse effects is needed. Patient's subjective experience of medication should be given more consideration. Total discontinuation rates provide a useful global outcome measure that incorporates tolerability and efficacy as well as patient and clinician viewpoints. Patients should be informed of common side-effects prior to treatment and monitored for their occurrence during treatment.

Declaration of interest RMH. has received honoraria from several pharmaceutical companies.

Knowledge of how the prevalence and severity of adverse effects vary for different antipsychotics allows clinicians to reduce the occurrence of these effects. We review the range of adverse effects associated with antipsychotics and their clinical impact, and give an overview of the various sources of data on adverse effects and their relative strengths and weaknesses. Potential problems in interpreting the evidence base are considered and the importance of the patients' perspective emphasised. We conclude with an examination of total discontinuation rates as a global measure of effectiveness that incorporates both tolerability and efficacy.

RANGE AND CLINICAL IMPACT OF ADVERSE EFFECTS

Antipsychotics are associated with a wide range of potential adverse effects (Appendix 1) which can affect the patient in several ways (Fig. 1). For example the stiffness, slowness of movement and tremor of antipsychotic-induced parkinsonism (Dursun et al, 2004) can make it difficult for a patient to write, fasten buttons and tie shoelaces, leading to reduced quality of life. The blank 'mask-like' expression, tremor, stooped posture, drooling and abnormalities of gait (including lack of arm swing) are easily observable by others and mark the patient out as 'different', hence contributing to stigma. When severe the festinant gait may result in falls and injury, particularly hip fracture in older patients. Patients who recognise the link between these problems and antipsychotic medication may miss out doses or stop medication totally.

Many patients who adhere poorly to medication do not inform their clinical team of this and some go to great lengths to hide their non-adherence (covert non-adherence). Poor adherence during acute treatment of psychosis leads to chronic symptoms whereas poor adherence after remission increases the risk of relapse. Both may have serious consequences, including self-harm, aggression and readmission to hospital. When clinician and patient are aware of adverse effects, treatment can be adjusted to minimise the problems (e.g. dose reduction of the antipsychotic, prescription of an anti-Parkinsonian agent or a switch to an alternative antipsychotic with less propensity to cause the adverse effect).

SOURCES OF DATA

Data on adverse effects are available from a range of sources. These include randomised controlled trials (RCTs), naturalistic studies, part-marketing surveillance, and non-randomised and open trials. Open and non-randomised trials are methodologically inferior to double-blind randomised controlled trials but nevertheless contribute to the evidence base. All data sources can be considered as being pieces of a jigsaw; the full picture of drug tolerability is only evident when all the pieces are put together.

Randomised controlled trials

Strengths Double-blind randomised trials are regarded as the gold standard level of evidence for the following reasons.

(a) Randomisation reduces the risk of bias in baseline characteristics and so makes it more probable that differences in outcome reflect differences between the treatments being investigated.

(b) Comparative data can be obtained against either placebo or one or more comparator drugs. Placebo data are particularly helpful in identifying the baseline rate of adverse effects independent of treatment with an active drug. Many potential adverse drug effects (e.g. weight gain, sexual dysfunction, onset of diabetes) are multifactorial and occur in the general population.

(c) Prospective assessment allows accurate measurement of adverse effects. This may involve the use of standardised rating scales (Table 1).

(d) Patient and rater bias are eliminated by masking.

In practice these advantages are not always as clear-cut as they seem. For example, relatively few trials assess the success of masking and when they do the methods
used, analysis and reporting of the results are inconsistent (Boutron et al, 2005). A review of papers claiming to be RCTs, published in the British Journal of Psychiatry and the American Journal of Psychiatry, showed that reporting of the method of randomisation was uncommon (Ogunsipe et al, 1999). The authors concluded that the RCT status of some of the papers must therefore be in doubt.

Weaknesses

Although RCTs can allow accurate information on the incidence and prevalence of adverse effects to be gathered, most trials of antipsychotics have relatively small samples and are short term, lasting 4-8 weeks. Such studies may underestimate early-onset side-effects that are uncommon and cannot provide data on side-effects that develop in the medium and long term. For example, amenorrhoea is an adverse effect of antipsychotics that reflects hyperprolactinaemia (Wieck & Haddad, 2003). In the Schizophrenia Outpatient Health Outcome (SOHO) study the baseline prevalence was approximately 33% of women (Haro & Salvador-Carulla, 2006). Definitions of amenorrhoea differ; if it is defined as three consecutive missed episodes of menstruation then it will be impossible to detect in a drug trial of less than 12 weeks’ duration. The inability of short-term trials to provide data on long-term tolerability, including weight gain, sexual functioning and metabolic parameters, is a major weakness, as in clinical practice antipsychotics are often prescribed to patients for several years or even decades. This drawback has been partly addressed by two recently published RCTs with relatively long follow-up periods: the Cost-Utility of the Latest Antipsychotic Drugs in Schizophrenia study (CUtLASS) in the UK (Jones et al, 2006), which followed patients for 1 year, and phase 1 of the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study in the USA (Lieberman et al, 2005), which followed patients for 18 months. Nevertheless neither study is long enough to accurately assess the risk of tardive dyskinesia.

Prospective studies of conventional antipsychotics indicate a cumulative incidence of tardive dyskinesia of approximately 20% over 5 years of treatment (Morgenstern & Glazer 1993).

The protocols of most RCTs exclude patients with significant comorbid medical conditions. Consequently the tolerability of drugs in people with physical illness (for example those with hepatic and renal impairment) is often unknown prior to licensing. Some trials may also underestimate tolerability because there may be incentives for patients to remain in the trial that do not operate in clinical practice.

Naturalistic studies

Naturalistic studies, including pharmaco-epidemiological studies, have the advantage of assessing ‘real world’ patients. Pharmaco-epidemiological studies can have very large samples, enabling relatively rare adverse effects to be investigated. Both incidence and prevalence data can be generated. These studies are limited to data recorded on computerised record systems and the absence of relevant data may prevent adjustment for potential confounding factors. Furthermore, the lack of randomisation limits attribution of causality. Data regarding the safety of drugs in pregnancy derive from post-marketing surveillance and naturalistic studies because pregnant women are invariably excluded from RCTs.

Post-marketing surveillance

Post-marketing surveillance is an essential component of assessing drug safety and

Table 1 Examples of ratings scales used to assess side-effects of antipsychotics

<table>
<thead>
<tr>
<th>Side-effect</th>
<th>Scale(s)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tardive dyskinesia</td>
<td>Abnormal Involuntary Movements Scale (AIMS)</td>
<td>Guy (1976)</td>
</tr>
<tr>
<td>Akathisia</td>
<td>Barnes Akathisia Scale</td>
<td>Barnes (1989)</td>
</tr>
<tr>
<td></td>
<td>AMDP–S</td>
<td>Collegium Internationale Psychiatriae</td>
</tr>
<tr>
<td></td>
<td>Liverpool University Neuroleptic Side-Effect Rating Scale (LUNERS)</td>
<td>Scalarum (1986)</td>
</tr>
</tbody>
</table>

EPS, extrapyramidal side-effects; AMDP–S, Association for Methodology and Documentation in Psychiatry Adverse Event Questionnaire.
tolerability, and often provides the first evidence of adverse effects that are rare or confined to particular at-risk groups. Remoxipride was an antipsychotic marketed in the late 1980s. Trials indicated similar efficacy to haloperidol for treating positive and negative symptoms but with less risk of extrapyramidal side-effects. Following its introduction in Europe a significant number of cases of aplastic anaemia were reported (as many as 1 in 10,000). Remoxipride was withdrawn in 1993 (Fung et al., 2001). Pimozide is a conventional antipsychotic. Between 1971 and 1995, 16 deaths and 24 cases of serious cardiac events were reported to the Committee for the Safety of Medicines. This led to the following recommendations: (a) patients prescribed pimozide should undergo a baseline electrocardiogram (ECG) followed by annual ECGs; (b) if the QTc interval is prolonged, treatment needs to be closely supervised or withdrawn; and (c) pimozide should not be prescribed in conjunction with other drugs that prolong the QTc interval (Haddad & Anderson, 2002).

Post-marketing surveillance includes prescription event monitoring (Mann, 1998) and reports of adverse drug reactions (Gough, 2005). Various national and international regulatory bodies provide systems for post-marketing surveillance, an example being the UK yellow card system for reporting adverse drug reactions. Post-marketing surveillance is also conducted by pharmaceutical companies or by independent research companies employed by them. The potential conflict of interest inherent in manufacturers collecting, evaluating and reporting post-marketing data on their own products has been the subject of recent discussion (Fontanarosa et al., 2004). This point apart, post-marketing surveillance has several weaknesses: it relies on voluntary participation; underreporting is widespread; submitted reports may be of poor quality with inadequate detail; and the ability to confirm causality is limited. Incomplete numerator data on events and unreliable denominator data make it difficult to calculate rates of adverse events.

The withdrawal of drugs for safety reasons demonstrates that licensing is not a guarantee of safety and highlights the importance of the continuing assessment of tolerability and safety from further studies and post-marketing surveillance. Between 1960 and 1999 121 prescription drugs were withdrawn from worldwide markets for safety reasons (Fung et al., 2001). Drugs that act on the central nervous system were the most common category withdrawn; in a more detailed breakdown by drug class antidepressants were ranked fifth (7.4%). The top safety reasons for withdrawal among all drugs were hepatic (26.2%), haematological (10.5%), cardiovascular (8.7%), dermatological (6.3%) and carcinogenic issues (6.3%). The median time on the market for products where this information was available was 5.4 years, with approximately one-third being withdrawn within the first 2 years of initial marketing.

PROBLEMS IN INTERPRETING TOLERABILITY DATA

The researcher or clinician is faced with several problems when evaluating the literature on adverse effects of antipsychotics (Appendix 2).

**Limited data**

Many papers that report RCTs of antipsychotics provide little data on adverse effects and concentrate on efficacy. Where such data are provided they are often limited, for example until recently most trials of antipsychotics did not include any measures of glucose and lipid regulation.

**Drug carry-over effects**

A second problem is that most trials evaluate patients with chronic psychosis who must discontinue a previous antipsychotic before starting the trial. This makes drug carry-over effects inevitable. For example, the potential for weight gain associated with a particular antipsychotic is underestimated, as patients are likely to have gained weight during previous antipsychotic treatment, thus minimising their potential for further weight gain (Haddad, 2005). Assessing patients with first-onset psychosis who are drug naive overcomes this problem, but enrolling such patients into trials is notoriously difficult and such RCTs are rare.

**Bias in trial design**

Industry-sponsored trials are more likely to report results that favour the sponsor’s compound than are independent studies (Ahmer et al., 2005). Possible explanations include publication bias and bias in trial design. An example of the latter is that most RCTs of atypical antipsychotics employ haloperidol as the active comparator. Among the conventional antipsychotics, haloperidol is associated with a high incidence of extrapyramidal side-effects (EPS) and so it is not surprising that these studies generally report an advantage in relation to EPS for the atypical agents, an advantage that remains in meta-analyses (Geddes et al., 2000; Bagnall et al., 2003). In contrast, RCTs that have a low-potency conventional antipsychotic as comparator show no significant difference in the incidence of EPS for atypical antipsychotics other than clozapine (Leucht et al., 2003; Lieberman et al., 2005).

**Comparison between trials**

It is often necessary to compare data on adverse effects between trials. For example, the relatively few head-to-head RCTs of atypical antipsychotics make cross-study comparisons, despite their methodological pitfalls, a necessity. Furthermore, as estimates of the prevalence/severity of an adverse effect for any given drug will vary between trials, an adjusted value is often required. Meta-analysis is commonly used to allow data from different studies to be pooled and compared, but this approach is often not possible when analysing data on adverse effects because of varying methodologies used to assess such effects. For example, there are several scales to measure sexual function (Table 1). Parkinsonian symptoms are usually assessed using the Simpson–Angus Scale (Simpson & Angus, 1970), but some studies report the proportion of patients prescribed an anticholinergic drug, a clinical proxy for parkinsonism. Even when the same rating scale or measure is used, the outcome may be expressed in different ways. Parkinsonian symptoms may be reported as mean change in score on the Simpson–Angus Scale from baseline to end-point or as the number of patients with scores above a specified cut-off. Similarly, measures of weight change during a study include mean change in kilograms, the percentage of patients with increments of weight change (e.g. 0–5 kg, 5–10 kg, etc.) and the number of patients with an arbitrary measure of significant weight gain, (e.g. an increase of more than 7% of baseline weight).

**Outcome measures that lack clinical utility**

Many studies present data on adverse effects in terms of the mean change in an outcome measure (e.g. a rating scale or the blood
symptoms rather than determining their impact on patients. Recently there has been increasing interest in the subjective view of patients to treatment, including antipsychotic medication (Voruganti et al., 2000; Angermeyer et al., 2001; Hasler et al., 2004). There are several overlapping domains, including subjective satisfaction with treatment, subjective quality of life and subjective response to treatment.

Satisfaction with treatment
Patient satisfaction with treatment is influenced by multiple factors and not just symptom reduction (Hasler et al., 2004). Factors that predicted dissatisfaction with care in a large European study included unemployment, more severe psychopathology and a high rate of hospital admission (Thornicroft et al., 2004). Other reasons for dissatisfaction include lack of involvement in treatment planning or decision-making, lack of involvement with treatment options, drug side-effects and lack of information about these. In a UK survey of callers to a national mental health telephone helpline, distressing side-effects were strongly correlated with low treatment satisfaction (Fakhoury et al., 2001). In this survey the most distressing side-effects reported (percentage of patients with the side-effect who reported it as distressing) were weight gain (73%), depression (67%), insomnia (66%), difficulty thinking/concentrating (63%), sedation (59%) and sexual dysfunction (58%). Men were more likely to report sexual dysfunction as distressing and women more likely to report weight gain as distressing. Several studies indicate that adverse effects of antipsychotics are often not diagnosed or treated (e.g. Mitra & Haddad, 2007) and that psychiatrists tend to underestimate the distress that they cause (e.g. Day et al., 1998).

Subjective quality of life
Many factors influence a patient’s view of their quality of life, including positive and negative symptoms, depression, cognitive impairment, hospitalisation and perceived support (Thornicroft et al., 2004). Several studies have reported that quality of life is higher in patients treated with atypical antipsychotics than in those treated with conventional antipsychotics (Franz et al., 1997). However, in the CATIE study, the largest independent randomised double-blind study in schizophrenia research, there were no significant differences in psychosocial functioning (assessed using the Quality of Life Scale; Heinrichs et al., 1984) between those treated with atypical drugs and those treated with perphenazine, a conventional drug; all treatment groups showed modest improvement (Swartz et al., 2007). This is consistent with the CULASS study (Jones et al., 2006), which found no difference in quality of life scores between patients prescribed typical and atypical antipsychotics.

**Subjective response to treatment**
The Drug Attitude Inventory (DAI; Hogan et al., 1983) is an established tool that assesses acceptability and subjective tolerability (subjective response) of medication. Factors that influence subjective response include insight, previous experience of medication, health beliefs and the quality of the therapeutic relationship. In one study patients on atypical antipsychotics reported a more positive subjective response and a lower prevalence of dysphoria than those on typical antipsychotics (Voruganti et al., 2000). Subjective response, as assessed by DAI score, is strongly correlated with adherence (Awad & Hogan, 1994). However, adherence is influenced by many other factors, including the quality of the therapeutic relationship between the patient and physician or keyworker (Frank & Gunderson, 1990).

**Discontinuations owing to intolerability**
When interpreting trials there is often a tendency to consider individual side-effects in isolation (e.g. weight gain, EPS, hyperprolactinaemia, etc.) In reality patients often experience several adverse effects, and whereas each on its own may be minor together they may be a major burden. One measure of overall tolerability is the proportion of patients who stop treatment and cite side-effects as the cause. Although intolerability is a major cause of antipsychotic drug discontinuation in schizophrenia it often accounts for fewer discontinuations than lack of efficacy (Lieberman et al., 2005; Kinon et al., 2006; Haro et al., 2007). For example in phase I of the CATIE study patients were randomised double-blind to one of five antipsychotics and followed for up to 18 months. In four of the five drug cohorts more patients stopped treatment for lack of efficacy than for intolerability (Lieberman et al., 2005; Fig. 2). In the naturalistic SOHO
study the percentage of patients discontinuing treatment over 3 years because of lack of efficacy exceeded those discontinuing treatment for intolerability in all drug cohorts (Haro et al., 2007). This was also the case in a meta-analysis of four RCTs of olanzapine in schizophrenia (Kinon et al., 2006).

These results are consistent with a concept mapping study that investigated medication adherence in people with schizophrenia (Kikkert et al., 2006). Based on interviews with people with schizophrenia, carers and health professionals, ten clinically relevant clusters were identified that affected medication adherence. Medication efficacy was rated by patients and carers as the most important cluster affecting adherence, but professionals rated this as significantly less important, ranking it fifth out of the ten clusters. Conversely, patients and carers placed side-effects relatively low compared with positive aspects of medication, whereas professionals prioritised side-effects as the second most important cluster. So, compared with patients and carers, professionals overestimate the importance of adverse effects for adherence and underestimate the importance of efficacy.

**TOTAL DISCONTINUATION RATES: GLOBAL MEASURE OF EFFECTIVENESS**

Clinicians and patients need to balance adverse effects against the effectiveness of a drug in treating the psychiatric illness. If a patient obtains significant benefit from a drug they may be willing to put up with considerable adverse effects (as demonstrated with clozapine). Adverse effects are common with clozapine and regular monitoring of the full blood count is mandatory throughout treatment, owing to the risk of neutropaenia. Nevertheless, patients often accept the adverse effects, presumably because clozapine provides a level of symptom control for their treatment-resistant illness that was not achieved with previous antipsychotics. In problem-centred interviews with patients discharged from hospital on clozapine a wide range of side-effects were reported, including fatigue or sedation (56%), lack of motivation (21%), hyper-salivation (21%), anticholinergic effects (16%), weight gain (15%) and orthostatic hypotension (11%) (Angermeyer et al., 2001). Despite this nearly one-third of patients stated that they felt better as a result of clozapine and almost half expected a worsening of their mental state if they stopped the medication.

Thus it is helpful to have a global measure of the effectiveness of a drug that combines both tolerability and effectiveness in treating symptoms. One way to achieve this is to record the total discontinuation rate on the drug at a given time point or the time to discontinuation for any reason. As stopping medication in a trial is a joint decision made by patient and clinician, this outcome measure also has the advantage of incorporating the patient’s and the clinician’s views.

Discontinuation of treatment for any reason was the primary outcome measure in the CATIE study (Lieberman et al., 2005). The results of phase I of the study illustrate the importance of balancing efficacy and tolerability. Of the five antipsychotics in phase I, olanzapine was associated with the highest percentage of patients stopping treatment because of intolerability but the lowest percentage stopping treatment for lack of efficacy (Leberman et al., 2005). When discontinuations owing to lack of efficacy and intolerability were combined with discontinuations for other reasons then the total discontinuation rate for each of the five antipsychotics was lowest with olanzapine (Fig. 2). The high total discontinuation rates seen with all drugs in the CATIE study might partly reflect the double-blind design (Haddad & Dursun, 2006).

The total discontinuation rate has also been used as the outcome measure in several naturalistic studies (Hodgson et al., 2003; Tiilonen et al., 2006). Of particular note is the study by Tiilonen et al. (2006) in which a nationwide cohort of 2230 consecutive adults hospitalised in Finland for the first time with a diagnosis of schizophrenia or schizoaffective disorder were followed prospectively. Total rates of discontinuation, adjusted for the effect of confounders, were determined for the ten most commonly used antipsychotics and compared with haloperidol. Initial treatment with clozapine, perphenazine depot and olanzapine were associated with the lowest total discontinuation rates, and in all three cases these were significantly less than the rate associated with haloperidol. Significant differences were also seen between antipsychotics in the rates of readmission, with clozapine, perphenazine depot and olanzapine all being associated with significantly lower readmission rates than haloperidol.

**CONCLUSIONS**

Data on adverse effects are available from a range of sources, including randomised controlled trials, post-marketing surveillance and naturalistic studies. All sources of data carry their own advantages and disadvantages. The best overview of adverse effects comes from considering all sources together. There is inconsistent reporting of adverse effects across studies and many outcome measures lack clinical meaning. Future research would benefit greatly if standardisation for the reporting of adverse effects could be reached. The impact of side-effects on patients has not been sufficiently studied. It is important that the patient’s subjective experience, in which adverse effects have a role, are considered in the assessment of a drug. Although adverse effects are an important outcome, with many antipsychotics they account for less treatment discontinuation than lack of efficacy; this finding has been noted in naturalistic studies and in RCTs. Total...

![Fig. 2 Percentage of patients discontinuing medication for various reasons in phase I of CATIE study.](image)
discontinuation rates provide a useful global outcome measure that incorporates tolerability and efficacy and patient and clinician viewpoints. In clinical practice, patients should be informed of common side-effects prior to treatment and monitored for their occurrence during treatment.

REFERENCES


Appendix 1 Adverse effects of antipsychotics

<table>
<thead>
<tr>
<th>Antimuscarinic symptoms</th>
<th>Cardiovascular adverse effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry mouth</td>
<td>Ankles oedema</td>
</tr>
<tr>
<td>Blurred vision</td>
<td>Arrhythmias (in some cases related to QTc prolongation)</td>
</tr>
<tr>
<td>Cognitive impairment</td>
<td>Cardiomyopathy</td>
</tr>
<tr>
<td>Constipation</td>
<td>Myocarditis</td>
</tr>
<tr>
<td>Delirium</td>
<td>Postural hypotension</td>
</tr>
<tr>
<td>Urinary retention</td>
<td>Hyperprolactinaemia</td>
</tr>
<tr>
<td>Extrapyramidal syndromes</td>
<td>Decreased bone mineral density</td>
</tr>
<tr>
<td>Akathisia</td>
<td>Gynaecomastia</td>
</tr>
<tr>
<td>Dystonia</td>
<td>Menstrual irregularities</td>
</tr>
<tr>
<td>Parkinsonism</td>
<td>Sexual dysfunction</td>
</tr>
<tr>
<td>Tardive dyskinesia</td>
<td>Miscellaneous adverse effects</td>
</tr>
<tr>
<td>Metabolic effects</td>
<td>Blood dyscrasias</td>
</tr>
<tr>
<td>Hyperglycaemia and diabetes</td>
<td>Hyperosmolarization</td>
</tr>
<tr>
<td>Hyperlipidaemia</td>
<td>Neuroleptic malignant syndrome</td>
</tr>
<tr>
<td>Weight gain</td>
<td>Photosensitivity</td>
</tr>
<tr>
<td></td>
<td>Sedation</td>
</tr>
<tr>
<td></td>
<td>Seizures</td>
</tr>
<tr>
<td></td>
<td>Skin pigmentation</td>
</tr>
<tr>
<td></td>
<td>Thyroid abnormalities</td>
</tr>
</tbody>
</table>

Appendix 2 Potential problems in interpreting tolerability data

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited published data on adverse effects</td>
<td>Introduce regulatory requirement for assessment of certain adverse effects prior to licensing</td>
</tr>
<tr>
<td>Carry-over adverse effects from previous drug treatment</td>
<td>Promote importance of tolerability as an outcome measure</td>
</tr>
<tr>
<td>Bias in trial design</td>
<td>Incorporate drug-free run-in period in trial design</td>
</tr>
<tr>
<td>Different scales/measures used to rate the same side-effect in different trials</td>
<td>Study drug-naive patients</td>
</tr>
<tr>
<td>Outcome measures lack clinical utility</td>
<td>Less likely in non-industry sponsored studies</td>
</tr>
<tr>
<td>Trials vary markedly in terms of population characteristics, drug dosage and duration of treatment</td>
<td>Select appropriate comparator drug prescribed at an appropriate dose</td>
</tr>
<tr>
<td></td>
<td>Introduce standardisation for rating and reporting of adverse effects</td>
</tr>
<tr>
<td></td>
<td>Beware of adverse effect data that are reported in terms of the mean change from baseline to end-point on a rating scale or other dimensional measure</td>
</tr>
<tr>
<td></td>
<td>Consider whether outcome measure is clinically useful; would a categorical measure (e.g. the proportion of outliers) be more relevant?</td>
</tr>
<tr>
<td></td>
<td>Compare like with like when extrapolating trial data to an individual patient or when comparing trial data</td>
</tr>
</tbody>
</table>
The use of atypical antipsychotics in schizophrenia: lessons from CATIE.

Human Psychopharmacology (2006)

CNS Drugs Bulletin (2005)

Drugs in Schizophrenia Study (CUtLASS 1).

schizophrenia: Cost Utility of the Latest Antipsychotic Life of second- vs first-generation antipsychotic drugs in

Jones, P. B., Barnes, T. R., Davies, L., Jones, P. B., Barnes, T. R., Davies, L.,

Three-year antipsychotic effectiveness in the outpatient Haro, J. M., Suarez, D., Novick, D., Haro, J. M., Suarez, D., Novick, D.,

Patient satisfaction with outpatient psychiatric care of schizophrenia: observational versus randomized studies. European Neuropsychopharmacology, 17, 235–244.


Results of the Yale Tardive Dyskinesia Study. Archives of General Psychiatry, 50, 723–733.


Correspondence: Dr P. M. Haddad, Cromwell House, Cromwell Road, Eccles, Salford, Manchester M30 0GT, UK. Email: peter.haddad@bsstrht.nhs.uk.
Schizophrenia outcome measures in the wider international community

MOHAN ISAAC, PRABHAT CHAND and PRATIMA MURTHY

Background  Outcome of schizophrenia has been described as favourable in low- and middle-income countries. Recently, researchers have questioned these findings.

Aims  To examine the outcome studies carried out in different countries specifically looking at those from low- and middle-income countries.

Methods  Long-term course and outcome studies in schizophrenia were reviewed.

Results  A wide variety of outcome measures are used. The most frequent are clinical symptoms, hospitalisation and mortality (direct indicators), and social/occupational functioning, marriage, social support and burden of care (indirect indicators). Areas such as cognitive function, duration of untreated psychosis, quality of life and effect of medication have not been widely studied in low- and middle-income countries.

Conclusions  The outcome of schizophrenia appears to be better in low- and middle-income countries. A host of sociocultural factors have been cited as contributing to this but future research should aim to understand this better outcome. There is a need for more culture-specific instruments to measure outcomes.

Declaration of interest  None.

Schizophrenia may have a better outcome in low- and middle-income countries. The initial evidence for this came from the International Pilot Study of Schizophrenia (IPSS; World Health Organization, 1979) and was further strengthened by two subsequent studies, the Determinants of Outcome of Severe Mental Disorders (DoSMED; Jablensky et al, 1992) and the recently concluded International Study on Schizophrenia (ISoS; Harrison et al, 2001). A host of sociocultural factors have been cited as contributing to the better outcome in these countries, including lower expressed emotion, closely knit family structures and family interactions (Kulhara & Chakrabarti, 2001). However, there is little evidence for the beneficial influence of these factors. The World Health Organization (WHO) follow-up studies of the past 25 years have been unable to tease out the specific patterns and timing of cultural influences that determine a better prognosis in these countries. The strongly held belief in a better prognosis in these countries has been questioned in recent times (Patel et al, 2006).

The ISoS study, coordinated by the WHO, addressed the outcome and related issues in a 15- to 25-year follow-up of 14 culturally diverse schizophrenia cohorts. Although the outcome results were consistently more favourable from low- and middle-income countries, there was marked heterogeneity across the centres (Hopper & Vanderling, 2000). Removing Hong Kong left three centres in this category from India (one in Madras and two in Chandigarh). It might be helpful to examine which cultural aspects of the Indian subcontinent contribute to an improved outcome in people with schizophrenia (Patel et al, 2006).

Clinical symptoms

The Present State Examination–9 (PSE–9; Wing et al, 1974) has been used as the measure of clinical symptoms at baseline and during follow-up in almost all long-term studies from low- and middle-income countries. The PSE–9 assesses 140 symptoms grouped into 36 syndromes and measures the presence of symptoms in the previous month. In a 20-year longitudinal study from India (Thara, 2004), all syndromes registered decline, although slowness, loss of interest, concentration and simple depression registered an increase over the second 10 years whereas positive symptoms showed little difference. In this cohort only 5 out of 61 patients (8%) were continuously ill. Using a similar methodology, 10-year clinical outcome was reported as favourable in three-quarters of the sample (Thara & Eaton, 1996).

Outcome was classified into broad categories in the DoSMED cross-cultural study (Jablensky et al, 1992). The outcome criteria used were: good, remitting course with full remission; poor, continuous/ incomplete remission. A more recent long-term study from Singapore (Kua et al, 2003) described final outcome measures in similar broad domains—good, patient not receiving treatment, well and working; fair, patient not receiving treatment and not working, or receiving out-patient treatment and working; poor, patient receiving treatment and not working, or receiving in-patient treatment. This study included treatment, employment and hospitalisation as indicators of severity of clinical symptoms for patients with schizophrenia. Over two-thirds of patients had a good/fair outcome.

Measurement of positive or negative symptoms/syndromes has been used by most studies. However, neurocognitive symptoms were not properly covered in the outcome measures used. This domain...
has been receiving increased attention because of its association with functional recovery. Although there remains to be wide heterogeneity in cognitive functioning in individuals with schizophrenia, a number of recent studies from the West have suggested that cognitive deficits once established are relatively stable over time.

### Acute psychosis debate

Several researchers have argued that many patients with acute psychosis might have been included, contributing to good outcomes. Non-affective acute remitting psychoses are far more common in low- and middle-income countries (Susser & Wanderling, 1994). However, reanalysis of the data excluding patients with these psychoses did not change the results to any appreciable extent (Hopper & Wanderling, 2000). Indeed, such patients had slightly better outcomes in the high-income countries and excluding them increased the differences between high-income countries and India. Furthermore, other studies (Kulhara & Chandiramani, 1988) using more than one diagnostic definition and criteria for schizophrenia, also supported better outcome in low- and middle-income countries irrespective of diagnostic criteria.

### Duration of untreated psychosis

Studies from the West have shown that the duration of untreated psychosis (DUP) is associated with poorer outcome, with the relationship being strongest in the initial months of psychosis (Drake et al., 2000). This is particularly relevant in low- and middle-income countries where a significant number of patients come late to treatment. Reasons for this include lack of awareness, a strong belief in magical or religious causes, poor accessibility of healthcare systems and lack of community care (Isaac et al., 1981; Padnavathi et al., 1998). A cross-cultural study on pathways to psychiatric care (Gater et al., 1991) replicated these findings. Most patients are brought for treatment after a significant delay from the onset of symptoms.

### Table 1 Important outcome studies from low- and middle-income countries

<table>
<thead>
<tr>
<th>Source</th>
<th>Study/follow-up</th>
<th>Design</th>
<th>n at beginning/end</th>
<th>Instruments</th>
<th>Outcome, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thara et al (1994), India</td>
<td>Madras Longitudinal Study, 10-year follow-up</td>
<td>Prospective, three follow-up assessments at 2, 5 and 10 years</td>
<td>90/76</td>
<td>PSE, PPHS, IFS</td>
<td>Death, 11.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No symptoms, 65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Good, 72.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Psychotic, 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(7% continuously since inclusion)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Complete recovery, 14.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Complete remission with one or more relapses, 48.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Incomplete remission, 27.6</td>
</tr>
<tr>
<td>Thara (2004), India</td>
<td>Madras Longitudinal Study, 20-year follow-up</td>
<td>Prospective, four follow-up assessments at 2, 5, 10 and 20 years</td>
<td>90/61</td>
<td>PSE, PPHS, IFS (first 10 years), GAF</td>
<td>Asymptomatic at syndromal level, 59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Complete recovery, 8.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Continuously ill, 8.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Death, 17.7</td>
</tr>
<tr>
<td>Tirupati et al (2004), India</td>
<td>Untreated psychosis, 1-year follow-up</td>
<td>Prospective follow-up</td>
<td>49/49</td>
<td>PSE, PPHS</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Clinical, 29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Social, 35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Occupational, 51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Global, 31</td>
</tr>
<tr>
<td>Kurihara et al (2000, 2005), Bali</td>
<td>Treatment-naive schizophrenia, 5- and 11-year follow-up</td>
<td>Prospective</td>
<td>59/51, 5 years</td>
<td>PANSS, ESAS</td>
<td>Remission, 23.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Partial remission, 19.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Self-supportive, 39.1</td>
</tr>
<tr>
<td>Kua et al (2003), Singapore</td>
<td>20-year follow-up</td>
<td>Assessment at 5, 10, 15 and 20 years</td>
<td>402/216</td>
<td>Clinical interview on treatment and work status, no scales</td>
<td>Good/fair, 66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Working, 32</td>
</tr>
<tr>
<td>Lee et al (1998), Hong Kong</td>
<td>15-year follow-up</td>
<td>File review</td>
<td>100/70</td>
<td>PSE–9, DAS, LCS, PIRS-II, SFD, BRS</td>
<td>Recovered, 53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GAF score &gt; 60, 71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Working, 74</td>
</tr>
</tbody>
</table>

PSE: Present State Examination; PPHS: Psychiatric and Personal History Schedule; IFS: Interim Follow-Up Schedule; GAF: Global Assessment of Functioning; PANSS: Positive and Negative Syndrome Scale; ESAS: Emanuel Social Adjustment Scale; DAS: Disability Assessment Schedule; LCS: Life Chart Schedule; PFS: Psychological Impairment Rating Scale; SFD: Schedule for the Deceased; BRS: Broad Rating Schedule.

1. Excluding World Health Organization multicentre studies.
2. 5-year outcome strongly predicted long-term outcome; minority needed maintenance medications.
Out of 75 patients in India who were treatment naive and living with their family, 60% had a DUP of over 5 years and 36% over 12 years. Following treatment for 1 year, patients with a DUP of 5 years or less had shown good clinical outcome (Tirupati et al., 2004). All were treated with antipsychotics on an outpatient basis and none needed hospital admission. An encouraging observation was the notable treatment response despite many years of untreated illness. Short-term studies using score on the Positive and Negative Syndrome Scale (PANSS) as an outcome measure corroborated these findings (Philip et al., 2003).

The PSE–9, which is used in most studies, measures the presence of symptoms only for the past month, which is probably too brief for outcome assessment in a chronic illness such as schizophrenia. Moreover unanchored global judgements such as good, fair or poor are crude. One method for maximising specificity and generalisability is the use of structured instruments for interviews, defining core symptom variables with clearly outlined operational criteria and incorporating relevant existing scales with established psychometric credentials (McGlashan et al., 1988).

This has been reflected in studies such as ISoS (Harrison et al., 2001), in which the PSE–9 has been supplemented in most cases by the Schedule for the Assessment of Negative Symptoms (SANS; Andreasen, 1983), the Schedule for Clinical Assessment in Neuropsychiatry (SCAN; World Health Organization, 1998), and the Scale for Assessment of Positive Symptoms (SAPS; Andreasen, 1984) for outcome measurement with respect to clinical symptoms (Petersen et al., 2005).

**Hospitalisation/treatment-seeking**

Hospitalisation has been used as an outcome measure in several studies, generally as a proxy for acuteness of symptoms and functional disability (Burns, 2007, this supplement). In low- and middle-income countries, hospitalisation is more a reflection of policy and resource availability than an indication of need (Harrison et al., 2001). Many people with schizophrenia have never been treated or hospitalised, and assuming that they are asymptomatic or symptoms are not severe is unjustified (Isaac et al., 1981; Padmavathi et al., 1998). The lack of hospital beds and alternative systems of ‘residential care’ that exist in high-income countries limits the use of hospitalisation as a reliable outcome measure. According to the World Health Organization 2005 figures (http://globalatlas.who.int), the median number of hospital beds per 10,000 population in low- and middle-income countries is around 0.2 (India, 0.25) whereas it is 7 in high-income Western countries (UK, 6; Switzerland, 13.20).

Social factors such as unemployment in males, family awareness of the nature of illness and family type are strongly related to treatment-seeking in low- and middle-income countries (Srinivasan et al., 2001). Gender, level of literacy and economic status appear to be unrelated. Surprisingly, more florid positive symptoms (such as delusions, hallucinations or aggressive behaviour) were not associated with seeking treatment or hospitalisation. However, self-neglect seems to lead to treatment; an unhygienic, unkempt person was more noticeable in public or to visitors to the house, and family embarrassment stimulated treatment-seeking for the patient.

The use of complementary medicines and consultations with traditional healers is widely acknowledged in low-income countries such as India (Raguram et al., 2002), but their impact, apart from a likely placebo effect, has not been adequately studied.

**Mortality**

The mortality rate is often neglected in outcome studies, but recently high mortality rates have been reported from low- and middle-income countries (Patel et al., 2006). In schizophrenia outcome studies spanning 15–25 years, the proportion of patients who died or were lost to follow-up ranged from 23% in Chennai to over 50% in Chandigarh and Agra (Harrison et al., 2001). Thara’s (2004) study found a mortality rate of 10% at 10 years, which increased to 17% at 20 years. The mean age at death was 34.2 years, which is well below the national average life span of 60.5 years. A much higher mortality of 47% was reported in a 15-year follow-up study of patients from North India with early poor outcome. Out of 15 patients with a poor course of schizophrenia during the first 2 years, 7 had died before completion of follow-up (Mojtabai et al., 2001). A high mortality rate of over 10% has been reported from Ethiopia during follow-up periods of 1–4 years (Kebede et al., 2005). Suicide accounted for nearly half of the deaths of those under 35 years.

**Social functioning**

In low- and middle-income countries schizophrenia has been shown to have better outcome in terms of social and occupational functioning; social functioning more than clinical status influences the functional competence of people with schizophrenia (Verghese et al., 1990; Harrison et al., 2001).

The development of measures for the assessment of impaired social functioning lags behind clinical rating. This relative neglect of a standardised assessment of social adaptation may reflect an assumption that symptomatology is closely tied to impairment in social functioning. This pattern is repeated in research in low- and middle-income countries. Most social outcome measures are derived from scales measuring psychopathology (i.e. the PSE) or from the course of the illness (e.g. the Psychiatric and Personal History Schedule, PPFS; Verghese et al., 1983) (Srinivasan et al., 2001; Thara, 2004). Social functioning outcomes that were measured from PSE items could not distinguish social impairment from prevailing neurotic or psychotic conditions. The PPFS rates the availability and frequency of a patient’s social contact during 1 month preceding evaluation. These items refer to living in a household, close friends, casual friends and the presence or absence of social activity groups. Few studies have used locally derived scales (e.g. Eguma’s Social Adjustment Scale; Kurihara et al., 2000, 2005) to assess social and vocational outcome.

Several measures have been developed and validated for use in these populations. Scales from low- and middle-income countries include the Schedule for the Assessment of Psychiatric Disability (SAPD; Thara et al., 1988) and the SCARF Social Functioning Index (Padmavathi et al., 1995); and from high-income countries the Groningen Social Disability Schedule (GSDS; Wiersma et al., 1988), the Life Skill Profile (LSP; Rosen et al., 1989), the Social Function Scale (SFS; Birchwood et al., 1990), the Social Adaptive Functioning Evaluation (SAFE; Harvey et al., 1997), and the Independent Living Scale Survey (ILSS; Wallace et al., 2000). For most cross-cultural studies the Disability Assessment Schedule (World Health Organization, 2000) and
Global Assessment of Functioning (GAF; American Psychiatric Association, 1987) have been used.

Some measures of social functioning include items reflecting clinical symptoms which need to be distinguished from those of functioning. In the West, many patients reside in assisted living facilities whereas in low- and middle-income countries the majority live in the community and are cared for by family members. This important component has not been adequately represented in instruments for assessment of social functioning (Sarswat et al., 2006).

Recently the Social Occupational Functioning Scale (SOF) has been developed and validated in India, and has been found suitable for use in multiple settings such as out-patient clinics, facilities and rehabilitation centres (Sarswat et al., 2006). A younger age at onset but not gender was associated with greater impairment in social functioning. Although none of the items was related to overall psychopathology, the item scores were correlated with positive and negative symptoms (Sarswat et al., 2006).

A neurocognitive study from India showed a lack of association of cognitive deficits with social functioning, employment and work performance (Srinivasan & Tirupati, 2005). At the same time there is an association of negative symptoms with these parameters. Measures of social functioning (i.e. communication and interest) are strongly associated with work functioning.

Currently there are few studies using social functional outcome measures from low- and middle-income countries. Social functioning is an important domain and, although sometimes cumbersome to measure, urgently needs to be incorporated as a regular outcome measure.

**Employment**

People with schizophrenia in low- and middle-income countries are more likely to be employed than their Western counterparts. Srinivasan & Thara (1997) found an annual rate of employment of 63–73% in the first 10 years of follow-up in a cohort of 90 people with first-episode schizophrenia. Moreover, among untreated Indian people with schizophrenia almost one-third were employed (Padmavathi et al., 1998). Moreover, almost half obtained employment within a year of starting treatment with antipsychotics (Srinivasan et al., 2001). Generally, high employment rates (up to 75%) have been found in India (Thara, 2004). A similar trend is described among Chinese patients; nearly half were able to work after 5, 10 and 15 years of follow-up (Tsui & Wong, 1991). These rates of employment are markedly higher than those in similar populations in high-income countries (Mueser et al., 2001). The employment rate in the UK over the past 20 years among people with schizophrenia ranges from 4 to 31%, with most Western studies reporting a rate of between 10 and 20% (Marwaha & Johnson, 2004).

Workplace colleagues are found to be generally supportive in low- and middle-income countries (Srinivasan & Tirupati, 2005). They rarely make an issue of the unusual behaviour of the person with schizophrenia, in contrast to the West where a ‘hostile social climate’ may confront persons with schizophrenia, whose diagnosis denies them access to employment (Marwaha & Johnson, 2004). After treatment for an episode of illness their return to work is often accompanied by criticism and a denial of their skills.

This discrepancy derives in part from the easy availability of work in informal sectors, differences in socio-economic status and economic pressure owing to a lack of disability benefits in low- and middle-income countries. Employment is a critical factor for perceived recovery from illness in countries where families are reliant on the members for support. Future outcome studies need to incorporate in-depth analysis of these factors (i.e. type of job, sectors, performance, financial gain, absenteeism, etc.) to understand such significant variations.

**Marriage**

Marriage requires certain social abilities to be successful. In countries such as India, marriage is a once-in-a-lifetime event and is associated with a high degree of social approval. The sociocultural factors determining marriage and its maintenance are vastly different from those in Western societies.

Marital state can be considered an outcome measure, as its maintenance depends on stability and functioning of both partners. Schizophrenia manifests maximally at a marriageable age (i.e. around the 20s). Most studies from the West have reported low rates of marriage for people with schizophrenia (Nanko & Moridaria, 1993; Hutchinson et al., 1999). In contrast, a 10-year follow-up study from India found a high marital rate of 70% (Thara & Eaton, 1996). Good marital outcome in terms of marrying and maintaining the marriage was associated with good overall outcome in people with schizophrenia. Similarly, good marital outcome is related to a decrease in symptoms and a lower relapse rate.

There are conflicting reports on marital status and outcome as few follow-up studies have studied this in detail. A few studies found that being married favours a good outcome and others found no such relationship (Thara et al., 2003a). Outcome when schizophrenia develops after marriage, or in those who marry without disclosing their illness (which is common in low- and middle-income countries), needs further study.

Patients whose marriages have broken down, in addition to the stress of their mental illness, face hostility from family members and rejection by society. This can be a significant contributing factor to outcome in traditional societies (Thara et al., 2003a). In a qualitative study of 75 divorced/separated women with mental illness (57% of whom had developed their illness after marriage), Thara et al. (2003a) found that many did not get any maintenance from their husbands and were fully dependent on their parents for both social and financial security. They initially felt helpless and lost, but most ultimately reconciled themselves to their fate and were pessimistic about the future. However, few had contemplated suicide. Concerns of being a burden to their aged parents, and hostile criticism from parents and siblings further reinforced their plight. In the current era of rapid globalisation, the effects of diminishing social support and the increasing prevalence of the nuclear family warrant close examination of the effects of these social changes on outcome.

**Social/family support**

Social support as a predictor of outcome in low- and middle-income countries has attracted considerable attention. Recent studies propose that supportive and favourable attitudes among family members and the community contribute to the improved outcomes (Kurihara et al., 2000, 2005). The mean time spent in hospital by people with schizophrenia is approximately a fifth in Bali compared with Tokyo (Kurihara et al., 2000). Studies from Asian countries showed that less than 10% were hospitalised during follow-up, suggesting high levels of family involvement in patient care.
It is suggested that social support is increased for both patients and caregivers from the extended family. This minimises the damaging effects of the illness and improves outcome.

Migration, urbanisation, changes in family structure and social support networks, plus the increase in economic insecurity and widening social inequalities which are evident in low- and middle-income countries will change the social support available for people with schizophrenia and influence their outcome (Patel et al., 2006).

Illness beliefs

Research from low- and middle-income countries consistently shows that there is a significant delay in seeking treatment for people with schizophrenia. Misconceptions of illness, superstition and ignorance have been proposed as reasons. However, recent studies have shown that very few people still name supernatural factors alone as a cause of schizophrenia (Srinivasan & Thara, 2001). In a study of Indian patients (Srinivasan & Thara, 2001), supernatural cause was named by only 12% of families with a member with schizophrenia.

Burden of care

Although the overall burden of care might be comparable across cultures, there are different patterns reflecting different sociocultural factors. The issue is particularly relevant in low- and middle-income countries where the majority of patients stay with their caregivers. Pai & Kapur (1982) developed a semi-structured instrument covering six broad areas of burden (financial, family routine, leisure, interaction, effect on physical health and effect on mental health). They found that caregiver burden decreases with a reduction in the patient’s symptoms and improving drug adherence. Reduction of family burden is associated with better outcome and social functioning (Pai & Kapur, 1982). The Burden Assessment Schedule (BASS; Thara et al., 1988), which was developed and standardised in India, also indicated significant burden among caregivers, including inability to care for others, unpredictable behaviour of patients and dissatisfaction with the help from healthcare professionals (Thara et al., 2003b). Some family members have considered leaving their ill relatives in psychiatric hospitals for long-term institutionalisation.

Substance misuse

Comorbid substance misuse in schizophrenia has been described as a high-risk factor for poor outcomes, including treatment non-adherence, relapse, rehospitalisation, violence, victimisation, criminal justice involvement, HIV and hepatitis C (Swartz et al., 2006). Estimates of the prevalence of substance use disorders are up to 70%, depending on diagnostic assessment methods. Comorbid substance misuse (nicotine excluded) has been reported in about half of people with schizophrenia in the USA (Regier et al., 1990).

There are few epidemiological studies from low- and middle-income countries on the prevalence of substance misuse in the general population, and even fewer on prevalence among people with schizophrenia. A study in a psychiatric hospital showed that the prevalence of alcohol disorders among patients with severe mental disorders was much lower than in the general population (Carey et al., 2003). An out-patient study in Chennai showed that only 38% of males with schizophrenia were current smokers, which was not significantly different from the general population (Srinivasan & Thara, 2002). Srinivasan & Thara (2002) have argued that comorbidity with nicotine use is not entirely biological; ‘culture’ plays a major determinant role.

Substance misuse may thus be an important cultural factor among a host of others that may mediate the course and outcome of schizophrenia in low- and middle-income countries.

CONCLUSIONS

Low- and middle-income countries are characterised by a poorly organised healthcare sector, limited access to psychiatrists and longer duration of untreated psychosis. Yet the outcome of schizophrenia appears to be better. What has contributed to the better outcome in these countries is difficult to say. Prevailing cultural factors and the nature of care and support might in part contribute to outcome. Factors such as the role of family and caregivers need further study as they play a vital and dynamic role in the care and rehabilitation of people with schizophrenia in low- and middle-income countries. The measures for diagnosis and assessment of outcome used in Western countries might not be suitable. A simple example is that translations of measuring instruments are often unreliable. For any multicentre research, money, manpower and technical expertise are essential, and these have always been scarce in low- and middle-income countries, especially for mental health. This is one of the reasons for the few studies on outcome in these countries. In the majority of these studies, the various outcomes were not studied using standardised and culturally appropriate instruments. Longitudinal studies using parameters such as neurocognitive function and quality of life are almost non-existent. Most studies are hospital based and there is a need for well-designed community-based outcome studies in these countries. India, like many other low- and middle-income countries, represents a society in transition. Whether the current sociocultural patterns associated with good outcome will themselves change and in turn alter the outcome of schizophrenia needs to be examined through prospective studies.

APPENDIX

Factors apparently contributing to good prognosis of schizophrenia in low- and middle-income countries

Established
Less expressed emotion
Good social support
Tolerance of odd behaviour by society and family
Marriage

Doubtful
Less industrialisation and urbanisation
Early death of those with bad outcome
Increased prevalence of acute psychosis

Needs to be studied
Comorbid substance use
Duration of untreated psychosis
Pharmacological interventions

REFERENCES


Measurement of long-term outcomes in observational and randomised controlled trials

RICHARD HODGSON, CHRIS BUSHE and ROBERT HUNTER

Background  Randomised controlled trials (RCTs) are the gold standard for evaluating treatment efficacy. However, the outcomes of RCTs often lack clinical utility and usually do not address real-world effectiveness.

Aims  To review how traditional RCTs may be triangulated with other methodologies such as observational studies and pragmatic trials by highlighting recently reported studies, outcomes used and their respective merits.

Method  Literature review focusing on drug treatment.

Results  Recently reported observational and some pragmatic studies show a degree of consistency in reported results and use outcomes that have face validity for clinicians.

Conclusions  No single experimental paradigm or outcome provides the necessary data to optimise treatment of mental illness in the clinical setting.

Declaration of interest  R.H. and R.H. have received funding from several pharmaceutical companies. C.B. is an employee of Eli Lilly UK. Funding detailed in Acknowledgements.

Evaluating treatment outcomes in mental illness presents unique and formidable challenges. The natural course of many psychiatric disorders is cyclical with spontaneous remission a distinct possibility (Ciompi, 1980). Environmental factors are important but poorly understood. Mental illness continues to be characterised in terms of symptoms despite advances in understanding pathogenesis. Currently, most published pharmacotherapy clinical trial data derive from trials performed to prove efficacy and safety to regulatory authorities. Thus clinicians making treatment decisions are commonly presented with a series of randomised controlled trials (RCTs) undertaken to meet regulatory requirements, with outcomes that are neither pragmatic nor easily transferable to clinical practice.

It is assumed that psychiatrists will base their treatment on the best available evidence but what is the best available evidence for a given clinician? Many factors are relevant and include personal experience, the literature, anecdote, opinion leaders, the pharmaceutical industry, guidelines and cost. However, little is known about actual prescribing and other treatment decisions (Hoblyn et al, 2006). Clinicians, purchasers and user advocates are also demanding more pragmatic end-points, and longer trials have shown the utility of relapse rates, hospitalisation and discharge rates as outcome measures (Csernansky et al, 2002).

Thus in 2007 ‘best available evidence’ is generally accepted as the RCT, but the available RCT evidence is at best incomplete, and at worst, flawed (Black, 1996). The aim of this paper is to show practising clinicians the spectrum of quantitative evidence and pragmatic outcomes.

EVOLUTION OF CLINICAL TRIALS

Since the 1940s the RCT has been the principal method of comparing the efficacy of all forms of medical treatment, and the basic concept has been developed and refined to further reduce bias. This has been evident in psychiatry with the development of rating scales and classification systems which enhance reliability, if not always validity. The RCT has informed the development of evidence-based medicine, meta-analysis and the Cochrane Collaboration. Evidence-based medicine resulted in part from the realisation that clinical practice is often poorly informed by the best available evidence, and that many widely used treatments are either untested or have been shown to be ineffective (Lenzer, 2004). Evidence-based medicine has also been seen as a means by which policy makers, sometimes with academic support, control clinical freedom (Williams & Garner, 2002). Although RCTs have resulted in the discontinuation of fashionable but ineffective treatments such as insulin coma therapy (Ackner & Oldham, 1960), they are not without problems (Thornley & Adams, 1998). More recently other paradigms, including observational and pragmatic studies (Roland & Torgerson, 1998), have gained in acceptance and been recommended as having a useful role in evaluation of treatment by the National Institute for Health and Clinical Excellence (National Institute for Clinical Excellence, 2002).

RANDOMISED CONTROLLED TRIALS

In general an RCT assesses efficacy — whether the treatment works in a controlled environment — not whether it works in the real world (effectiveness) (Table 1). Many factors affect the relationship between efficacy and effectiveness. This is acknowledged in the CONSORT criteria for RCTs by the need to assess the generalisability of the results, although a framework for assessing and reporting this is lacking (Bonell et al, 2006). Trials have been criticised for not adhering to CONSORT guidelines, but even apparent adherence can lead to challenges (El-Sayeh et al, 2006).

Patient recruitment and selection bias

Whether clinically significant selection bias occurs during recruitment to clinical trials is contentious. Although Burns (2006) reported that the basic demography of patients in a large naturalistic study was...
Comparison of key features of randomised controlled trials and observational studies

<table>
<thead>
<tr>
<th>Modest numbers of patients</th>
<th>Large number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modest duration</td>
<td>Longer duration</td>
</tr>
<tr>
<td>High drop-out rate</td>
<td>Lower drop-out rate</td>
</tr>
<tr>
<td>Statistically significant results</td>
<td>Clinically meaningful results</td>
</tr>
<tr>
<td>Structured dosing regimen</td>
<td>Naturally selected dosing</td>
</tr>
<tr>
<td>Randomisation</td>
<td>Naturalistic treatment selection</td>
</tr>
<tr>
<td>Maximises internal validity</td>
<td>Maximises external validity</td>
</tr>
<tr>
<td>Minimal bias and variability</td>
<td>Generalisability</td>
</tr>
<tr>
<td>Homogeneous patient population</td>
<td>Heterogeneous patient population</td>
</tr>
<tr>
<td>Artificial adherence and population</td>
<td>Adherence not mandated, 'real' patients</td>
</tr>
<tr>
<td>Demonstrates efficacy</td>
<td>Assesses effectiveness</td>
</tr>
<tr>
<td>Excludes confounding treatments</td>
<td>Concomitant treatments allowed</td>
</tr>
<tr>
<td>Complex applied scales</td>
<td>Outcomes used in everyday clinical practice</td>
</tr>
<tr>
<td>Outcomes generally symptom focused</td>
<td>Outcomes include cost, adherence, resource use</td>
</tr>
</tbody>
</table>

similar to that of a widely reported RCT, other authors have noted that the more chaotic patient who is difficult to manage will not be entered into a clinical trial as, even if they consent, they will undoubtedly drop out of follow-up (Lester & Wilson, 1999; Harrison-Read et al, 2002). Trials rarely report the number of patients considered or screened for a trial who are never included. Although this is a CONSORT requirement, clinicians will make pre-screening decisions regarding eligibility that are never reported. This is a potential source of bias and might limit extrapolation of results. It is likely that these difficulties are a serious unreported bias in published RCTs for psychological treatments. For example, reviews of the impact of day hospital treatment have failed to take entry criteria into account, leading to potentially erroneous conclusions (Thornicroft & Strathdee, 1994). The need for informed consent might inadvertently affect the generalisability of data from RCTs. All trials of intramuscular olanzapine (Meehan et al, 2001; Wright et al, 2001) were conducted in patients who gave informed consent and, although positive, the results cannot be interpreted as indicating that the drug will be as effective in patients who are highly disturbed.

Although biases are reduced in RCTs they are not eliminated, and indeed specific biases may even be created. Aside from the increased practical difficulties of including older adults in clinical trials, only 4.2% of older patients with major depression meet the increasingly rigorous inclusion and exclusion criteria of phase 3 studies (Yastrubetskaya et al, 1997). Women have sometimes been underrepresented in RCTs primarily because of concerns regarding conception while on trial medication, although this may be changing.

Patients with comorbid disorders are usually excluded from RCTs and this does not allow trials to reflect the rate of substance misuse and physical ill health in people with mental illness (Phelan et al, 2001). Previous exposure to trial medication is often unreported, but McQuade et al (2004) reported that 25% of patients in this randomised trial had prior exposure to one of the evaluated drugs. Generally, RCTs do not control for previous number of admissions or other markers of 'difficult to treat' patients (Hodgson et al, 2005). This might lead to newer treatments being tried in patients who are more difficult to treat, which may lead to suboptimal results for newer treatments (Davis et al, 2003).

Rating scale outcomes

The outcome measures used in RCTs affect the generalisability of the results. Although these outcome measures have been refined over decades to improve reliability, in studies their use may affect the face validity of the results. Clinicians would have difficulties in understanding what a fall of 20% in score on the Positive and Negative Syndromes Scale (PANSS; von Knorring & Lindstrom, 1995) means in clinical practice. Indeed Kane et al (1988) suggested this as an outcome only for treatment-resistant patients and a recent analysis (Leucht et al, 2005) has shown that a drop of 50% in PANSS score may better equate to a Clinical Global Impression Scale (CGI, Haro et al, 2003) rating of 'much improved'.

Pragmatic outcomes

Rating scales might not reflect clinical reality and there may be dissonance between rating scale response and a pragmatic clinical end-point such as discharge from hospital (McCue et al, 2006). Pragmatic research and outcomes focus on whether an intervention works under real-life conditions and whether it works in terms that matter to the patient. However, if broader concepts are used, such as remission, relapse or rehospitalisation, then other problems emerge. Rehospitalisation is easily measured, but in an individual trial may be mediated by other variables such as admission criteria. Remission or response rates might have more clinical utility but have been criticised on the grounds of variability of results if an arbitrary cut-off is used, although sensitivity analysis can be used to assess the effect of changing parameters (Linden et al, 2006; van Os et al, 2006).

Rates of discontinuation of treatment may be a proxy for treatment effectiveness (Hodgson, 2005; Lieberman et al, 2005; Kinon et al, 2006). Kinon et al (2006) undertook a meta-analysis of RCTs of atypical antipsychotics using reported discontinuation as an outcome and found far more variability between drugs than might have been anticipated from the headline results, which usually (marginally) favour the sponsor’s product (Heres et al, 2006). Further exploration of these pragmatic end-points in long-term studies facilitate a better understanding of the face and predictive validity of rating scales. Any dissonance between comparator drugs using varied end-points might be cause for concern. A recent non-inferiority RCT comparing two atypical antipsychotics at 1 year showed consistency of superiority for one in parameters ranging from PANSS score to discontinuation and hospitalisation rates (www.clinicaltrialresults.org/drugdetails/?drug_name_id=187&sort-c.company_name&page=1&drug_id=509). However, use of outcomes such as hospitalisation might preclude cross-service comparisons. Quality of life has also been used as an outcome but although such measures are laudable, in practice the outcomes are difficult to measure and may not be amenable to change (Boardman et al, 1999).
Tolerability
Published RCTs have been criticised for inadequate reporting of side-effects and adverse events (Ioannidis & Lau, 2001; Papanikolaou et al, 2004). The incidence is usually reported but duration and severity are not. These are important variables and may make the difference between persevering with medication or abandoning a therapeutic trial. For data such as prolactin levels RCTs often report mean cohort values rather than pragmatically useful categorical rates (Bushe & Shaw, 2007).

Study length and drop out
Typically patients in secondary services receive treatment for periods of time that far exceed those of RCTs, which are often as short as 4 weeks. The Schizophrenia Outpatient Health Outcomes (SOHO) study (Haro et al, 2006) demonstrated continued improvement over 3 years. Short RCTs will not assess all tolerability issues and whether improvement is maintained. However, RCTs are getting longer (Lieberman et al, 2003; McQuade et al, 2004). The corollary of longer study periods is lower follow-up rates and, paradoxically, high follow-up rates might be an indicator of a biased study population. Drop-out rates over 6 weeks are on average 35% and at 6 months can be around 72% (Leucht et al, 2003; McQuade et al, 2004), making interpretation of data complex.

Randomised controlled trials are designed to minimise bias and in creating this artificial environment treatment effects may be obviated. Although the true masking of many trials has been debated (Moncrieff, 1997), clinicians cannot intervene in trials in a timely or appropriate manner. Doses and visits are predetermined, as is the ability to respond to potential side-effects. These issues are relevant to the placebo arm, as often placebo group patients are receiving a psychoactive drug such as lorazepam (Meehan et al, 2001; Wright et al, 2001). Randomised controlled trials are often designed to fulfil regulatory requirements to obtain marketing authorisations for a new drug. There will be significant delays between study conception, recruitment, follow-up and publication of results. Clinicians often anticipate this with off-label prescribing (Hodgson & Belgamwar, 2006). The reality is that few RCTs are ever undertaken by pharmaceutical companies after launch. This is for many reasons, including the relatively short patent life. Thus, when such RCTs are performed there is often a perceived need for the data to be available quickly. Rarely are these trials long term.

Evolution of the RCT paradigm is seen in the CATIE trial (Lieberman et al, 2005; Table 2). In addition to traditional outcome measures, continuation on an antipsychotic was used as an outcome. Such an outcome should resonate with clinicians as medication is most commonly discontinued owing to lack of effectiveness or side-effects (Hodgson, 2005). Meta-analysis shows that lack of effectiveness is the major reason for discontinuation and differentiates between atypical antipsychotics in RCTs. In contrast, discontinuation for side-effects is relatively uniform (Kinon et al, 2006).

For the reasons above, RCTs fail to provide the clinician with all the necessary information to prescribe confidently. In order to prescribe a new product the clinician uses previous experience, critical review of early results and the experience of others. In other words the clinician is in effect, albeit informally, undertaking a naturalistic/observational study. The definition of an observational study can be problematic, but in the context of this paper we have identified the key element as a research design where the allocation of treatment is not fully under the control of the researcher (Table 1).

OBSERVATIONAL STUDIES

Limitations
There are notable long-term observational follow-up studies in psychiatry (Giompi, 1980; Harding, 1988) which illustrate the natural history of schizophrenia over decades. Given this expertise, it is perhaps surprising that there are so few studies looking at treatment effects over the longer term, especially as many potential outcome measures could be collected routinely. Observational studies have design faults that limit their interpretation (Table 1). Most importantly, true randomisation cannot occur in an observational study. However, the strengths of observational studies mirror the weaknesses of RCTs, and it is for this reason that National Institute for Health and Clinical Excellence (NICE) has argued for well-conducted observational studies to demonstrate effectiveness. Observational studies might also represent the only method for studying certain aspects of treatment when masking is not possible or ethical concerns preclude randomisation (Cook & Campbell, 1979). Indeed, in service evaluation studies randomisation may interfere with the dependent variable and observational studies often exploit service inequalities (Dean et al, 1993). Another potential bias in observational studies is rating bias, although the SOHO study has shown high correlations between clinician and patient ratings. With end-points such as hospitalisation, bias is minimised, especially if these data are collected routinely (Hodgson et al, 2001).

Observational studies have been criticised because they are believed to overestimate treatment effects. However, recent comparison between RCTs and observational studies does not support this view (Benson & Hartz, 2000; Concato et al, 2000; Kasper et al, 2001). Concato et al (2000) challenge the accepted hierarchy of clinical designs by reviewing outcomes from various methodologies in a variety of study areas and conclude that observational studies neither over- nor underestimate treatment effects to any significant degree. They opine that observational studies are more likely to produce homogeneous results as they include a broad spectrum of the population at risk. In addition, there is less chance of systematic treatment biases because of the broad treatment population.

Recent observational studies
The CATIE study (Lieberman et al, 2005), an RCT sponsored by the National Institute of Mental Health, compared the outcome of atypical antipsychotics with the typical antipsychotic perphenazine and also incorporated a switching strategy to evaluate clozapine. The results mirror those of Tiihonen et al (2006) in that clozapine and olanzapine were the only oral atypical antipsychotics to demonstrate lower discontinuation rates when compared with oral first-generation and other second-generation antipsychotics. The study reported by Tiihonen et al (2006) is particularly noteworthy as it follows a nationwide cohort of over 2000 people with first-episode schizophrenia for up to 7 years. In addition to showing differences in resorption and relapse rates between commonly available antipsychotics in Finland, it also shows the effectiveness of medication in reducing suicide and physical morbidity (adjusted relative risk 37.4, 95% CI 5.1–276 and 12.3, 95% CI 6.0–24.1 respectively). The relative therapeutic effects of the drugs studied did not vary...
whether discontinuation or rehospitalisation was considered, and this is echoed in the SOHO study (Haro et al, 2006). Another long-term study of over 500 patients in England (Hodgson et al, 2005) demonstrated the same rank order of effectiveness of oral atypicals using medication discontinuation as an outcome. In this study it was apparent that clozapine was being used for a treatment-resistant cohort. Taylor et al (2006) studied duration of treatment as a proxy in a Scottish population over 3 years and reported similar results to Tiihonen et al (2006) and Hodgson et al (2005).

McCue et al (2006) in a randomised open-label study of atypical antipsychotics and haloperidol in in-patients using the Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962) and time to discharge as outcome measures found similar effectiveness between haloperidol, olanzapine and risperidone and that these drugs were significantly better than aripiprazole and quetiapine. However, there was a dissonance between time to discharge and the BPRS outcomes, which might suggest that rating instruments are not sensitive to important changes that influence management, at least in the short term. Although haloperidol was equal to risperidone and olanzapine it was associated with more extrapyramidal side-effects. Jones et al (2006) failed to detect any differences in effectiveness between first- and second-generation antipsychotics and reported no difference in extrapyramidal-type side-effects, in stark contrast to many other RCTs. A recent RCT of 400 first-episode patients (McEvoy et al, 2006) compared olanzapine, quetiapine and risperidone over 1 year and failed to detect a difference in discontinuation rates between these drugs although olanzapine had a significantly greater effect on positive symptoms. Discontinuation was associated with poor response ($P<0.001$) and poor medication adherence ($P=0.02$).

In general, RCTs are powered for one primary outcome which does not always reflect primary clinical concern (McQuade et al, 2004). As observational studies are larger, there is more scope for legitimate subgroup analysis, such as treatment effect on those with comorbid disorder. The 3-year results of the SOHO study provide insights into social function and factors associated with relapse and remission. These are consonant with other independent studies and increase the face validity of this study. Although the SOHO study demonstrates relatively high switching rates for some medications, 65% of patients achieved remission, which resonates with the results of other long-term studies (Ciompi, 1980; Harding, 1988).

### Observational studies and safety

Although often not acknowledged as such, post-marketing surveillance is essentially an observational study, albeit often poorly conducted (Vray et al, 2005). However, post-marketing surveillance often reports important safety information that was not

---

**Table 2**  Key recent observational and pragmatic studies and randomised controlled trials in schizophrenia

<table>
<thead>
<tr>
<th>Reference</th>
<th>Methodology</th>
<th>Study size and follow-up</th>
<th>Setting</th>
<th>Key outcome measures</th>
<th>Key findings</th>
<th>Funding source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hodgson et al (2005)</td>
<td>Observational</td>
<td>502 patients up to 7 years</td>
<td>England</td>
<td>Medication discontinuation</td>
<td>Lowest discontinuation rate with clozapine, then olanzapine, then risperidone</td>
<td>Unrestricted grant from pharmaceutical industry</td>
</tr>
<tr>
<td>Haro et al (2006)</td>
<td>Observational</td>
<td>10 000 patients for 3 years</td>
<td>10 European countries</td>
<td>Medication discontinuation and remission</td>
<td>Lowest discontinuation rate and highest remission rate with clozapine, then olanzapine, then risperidone</td>
<td>Pharmaceutical industry</td>
</tr>
<tr>
<td>Taylor et al (2006)</td>
<td>Observational</td>
<td>958 patients for up to 3 years</td>
<td>Scotland</td>
<td>Duration of treatment</td>
<td>Duration of treatment longest with clozapine, then (in rank order) olanzapine, risperidone, amisulpiride and quetiapine</td>
<td>Independent</td>
</tr>
<tr>
<td>Tiihonen et al (2006)</td>
<td>Observational</td>
<td>2230 first-episode patients up to 7 years</td>
<td>Finland</td>
<td>Discontinuation and hospitalisation rates</td>
<td>Lowest relapse with oral medication for clozapine, then (in rank order) olanzapine, thiouracil, sertraline, risperidone and chlorpromazine</td>
<td>Government</td>
</tr>
<tr>
<td>Lieberman et al (2005)</td>
<td>RCT</td>
<td>1493 patients up to 18 months</td>
<td>USA</td>
<td>Medication discontinuation</td>
<td>Olanzapine most effective. No difference between other studies</td>
<td>Government</td>
</tr>
<tr>
<td>McEvoy et al (2006)</td>
<td>RCT</td>
<td>400 first-episode patients for 1 year</td>
<td>USA</td>
<td>Duration of treatment</td>
<td>No difference between olanzapine, quetiapine and risperidone</td>
<td>Pharmaceutical industry</td>
</tr>
<tr>
<td>McCue et al (2006)</td>
<td>Pragmatic</td>
<td>Hospitalised patients for at least 3 weeks</td>
<td>USA</td>
<td>Hospital discharge and BPRS</td>
<td>Haloperidol, olanzapine and risperidone more effective than aripiprazole, quetiapine and ziprasidone</td>
<td>Independent</td>
</tr>
</tbody>
</table>

**RCT**, randomised controlled trial; **BPRS**, Brief Psychiatric Rating Scale.
apparent from RCTs. The association between blood dyscrasias, clozapine and remoxipride are prime examples. In general, RCTs provide useful information on common adverse events, but identifying the relative risk of uncommon adverse events is realistically possible only in observational trials. In this regard, adverse event reporting in observational trials has been shown to enhance safety during the trial and facilitate the role of data monitoring committees and institutional review boards confronted with multiple reports of adverse events (Califf & Lee, 2001).

COMMON METHODOLOGICAL ISSUES

Analysis

Both RCTs and observational studies present difficulties in analysis. In RCTs high attrition rates have led to intention-to-treat analyses with a variety of statistical techniques evolving to accommodate these drop outs. These include last-observation-carried-forward (LOCF) analysis and mixed model repeated measures (MMRM); LOCF assumes that data are missing completely at random and that the patient’s condition would remain constant; both assumptions are unlikely; MMRM is valid under less restrictive assumptions with use of missing data dependent on other measured factors (Mallinckrodt et al, 2003).

Randomised controlled trials have highlighted relatively high switching rates between therapies and potentially confounding baseline variation, with lower rates measured in observational studies. Baseline variation can be accommodated in analysis but, as with drop out from RCTs, it cannot be assumed that this variation is random and may reflect clinical practice. For example, in the study reported by Hodgson et al, (2005) and the SOHO study (Haro et al, 2006) young men with multiple illness episodes were more likely to receive clozapine.

Switching treatments within an observational study can be studied using marginal structural models (MSM), a new class of causal models that allow for improved adjustment of confounding in longitudinal data analysis in naturalistic settings by consistently estimating the parameters of the inverse-probability-of-treatment weighted estimators (Mortimer et al, 2005); MSM are an extension of propensity scoring to longitudinal data. Whereas propensity scoring controls for selection bias by re-weighting observations to produce ‘balance’ between groups, MSM do the same but in a longitudinal fashion; MSM allow estimation of the causal effect of treatments in longitudinal naturalistic data when patients switch or stop treatment, even in the presence of missing (at random) data and time-varying confounding variables.

Patient concordance and sample size

In estimating treatment effects both RCTs and observational studies are challenged by patient concordance. Drug levels, which are highly variable for many psychotropics, are not routinely used, with pill counting being a common concordance measure in RCTs. However, poor adherence may underestimate treatment effects. Patient and clinician choice is important in determining outcome (Black, 1996) and controlling for these variables in RCTs limits the exploration of these factors. Zelen (1979) has advocated a methodology that has the advantage that, before providing consent, a patient will know whether an experimental treatment is to be used. Further development of patient and clinician preference trials has been described (Korn & Baumrind, 1991; Wennberg et al, 1993). McCue et al (2006) demonstrate that physician knowledge of a treatment might enhance optimum treatment dosing.

The nature of observational studies allows large sample sizes that add to the power of the study, facilitate subgroup analysis and provide data for robust sample size estimates for RCTs. Although in general appropriate sample sizes are important in RCTs, the superiority of those with larger sample sizes over those with smaller samples has been challenged with regard to overestimating treatment effects (Contopoulos-Ioannidis et al, 2005).

Publication bias and sponsorship

Publication bias might also affect the two methodologies. Given the hierarchy of evidence, journals may be less willing to accept observational studies (Barton, 2000). Journals are less likely to publish negative studies and both methodologies are potentially biased by the study sponsor, with positive results often being associated with the vested interest of the sponsor (Als-Nielsen et al, 2003). However, a review of atypical antipsychotic trials and funding sources indicates that this is not invariably so (Heres et al, 2006). Moreover, government-funded trials cannot be assumed to be unbiased (Coyne, 2006)

THE WAY FORWARD

The pre-eminence of RCTs and regulatory requirements has led to maintenance of the status quo in clinical drug trial development. Once a drug receives its marketing authorisation then further trial work is often aimed at developing markets rather than ascertaining whether the drug is effective. These concerns are just as relevant to psychotherapy and other non-pharmacological interventions. Making the trials as much like routine practice as possible may help to make RCTs more feasible and enhance external validity (so-called pragmatic trials; Hotopf, 2002). Although pragmatic trials may eschew some features of RCTs, such as double blinding, careful consideration may significantly reduce bias (Schulz et al, 1995). Patient recruitment is broad and may not be diagnostically driven (e.g. frequent attendees at a general practitioner surgery or people who self-harm). Outcomes, such as a reduction in suicide or episodes of violence, are clinically significant. Patient preference is an important variable in treatment choice which is negated in a traditional RCT, but patient preference trials have been reported (Ward et al, 2000) and may be particularly relevant when masking is not possible. The CATIE study (Lieberman et al, 2005) has many features of a pragmatic trial, such as narrow exclusion criteria and medication discontinuation as an outcome.

Randomised controlled trials and observational studies are not mutually exclusive, and there are examples from other areas of medicine of two designs running in parallel. For example, several studies quoted in Benson & Hartz (2000) in coronary artery disease illustrate the merits of enhancing an RCT by the addition of observational data from a concurrent registry of all non-randomised patients in the same centres. This approach improves the quality of observational research, since the same rigorous attention to detail in defining eligible patients, maintaining follow-up and recording outcomes is applied in both the randomised and the observational cohorts. The observational cohort may still suffer from selection bias, but there is a greater likelihood that its causes can be identified. The corollary also applies in that the
observational cohort inform on the typicality of the experimental group.

Rapid changes in methodology without bridging links with older methodologies may preclude legitimate comparison and subsequent meta-analysis. However, advances in the understanding of the biological and psychological mechanisms of mental illness will also dictate the evolution of relevant end-points. This is typified by the increasing interest in cognitive outcomes (Stroup et al., 2003) for which NICE recommends audits and provides standardised templates. This is another potential for supplementing treatment information and should facilitate the collection of data pools that inform treatment practice. The introduction of new treatment presents the possibility of mirror image studies (Hodgson et al., 2002) that allow some measure of utility, although regression towards the mean precludes overinterpretation of the results.

CONCLUSIONS

The RCT has served medicine well but evaluation of treatment needs reviewing for the 21st century. Outcomes need to be more clinically relevant and comparable with those from other trial methodologies. Biases in recruitment need to be addressed and post-marketing surveillance needs a more robust approach, as does monitoring of fidelity to treatment or service delivery models. In part this could be achieved with naturalistic studies, audits and mirror image studies. Without such additional information, treatments cannot be tailored effectively to the patient. Dogma should not be allowed to drive the experimental paradigm agenda as no current research design provides comprehensive clinical information.

ACKNOWLEDGEMENTS

R. Hunter has received funding from NHS Quality Improvement Scotland and the Chief Scientist Office, Edinburgh.

REFERENCES


LONGER-TERM OUTCOMES IN TRIALS
Assessing and interpreting treatment effects in Clinical antipsychotic trials of intervention et al, 2003


