Psychiatry in pictures


Highlights of this issue

Highlights of this issue
Sukhwinder S. Shergill

EDITORIALS

Patient choice in psychiatry
Chiara Samele, Simon Lawton-Smith, Lesley Warner, and Jeevi Mariathasan

Asylum claims and memory of trauma: sharing our knowledge
Jane Herlihy and Stuart W. Turner

REVIEW ARTICLES

Theory of mind in schizophrenia: Meta-analysis
Mirjam Sprong, Patricia Schothorst, Ellen Vos, Joop Hox, and Herman Van Engeland

PAPERS

Cost-effectiveness of first- v. second-generation antipsychotic drugs: results from a randomised controlled trial in schizophrenia responding poorly to previous therapy
L. M. Davies, S. Lewis, P. B. Jones, T. R. E. Barnes, F. Gaughran, K. Hayhurst, A. Markwick, H. Lloyd the CUtLASS team

Metabolic disease and cardiovascular risk in people treated with antipsychotics in the community
Paul Mackin, David Bishop, Helen Watkinson, Peter Gallagher, and I. Nicol Ferrier
Rates and correlates of employment in people with schizophrenia in the UK, France and Germany
Steven Marwaha, Sonia Johnson, Paul Bebbington, Mai Stafford, Matthias C. Angermeyer, Traolach Brugha, Jean-Michel Azorin, Reinhold Kilian, Karina Hansen, and Mondher Toumi

Determining when impairment constitutes incapacity for informed consent in schizophrenia research
Scott Y. H. Kim, Paul S. Appelbaum, Jeffrey Swan, T. Scott Stroup, Joseph P. McEvoy, Donald C. Goff, Dilip V. Jeste, J. Steven Lambert, Adrian Leibovici, and Eric D. Caine

Young people who self-harm
Robert Young, Michael Van Beinum, Helen Sweeting, and Patrick West

The Depression Scale as a screening instrument for a subsequent depressive episode in primary healthcare patients
Outi Poutanen, Anna-Maija Koivisto, Matti Joukamaa, Aino Mattila, and Raimo K. R. Salokangas

Heroin-assisted treatment for opioid dependence: Randomised controlled trial
Christian Haasen, Uwe Verthein, Peter Degkwitz, Juergen Berger, Michael Krausz, and Dieter Naber

Brain opioid receptor binding in early abstinence from opioid dependence: Positron emission tomography study
Tim M. Williams, Mark R. C. Daglish, Anne Lingford-Hughes, Lindsay G. Taylor, Alexander Hammers, David J. Brooks, Paul Grasby, Judith S. Myles, and David J. Nutt

Long-term outcomes after discharge from medium secure care: a cause for concern
Steffan Davies, Martin Clarke, Clive Hollin, and Conor Duggan

Impact of sexual violence on disclosure during Home Office interviews
Diana Bögner, Jane Herlihy, and Chris R. Brewin

SHORT REPORTS

Influence of drug company authorship and sponsorship on drug trial outcomes
Tongeji Tungaraza and Rob Poole

Depressive symptoms during pregnancy and low birth weight at term: Longitudinal study
Jonathan Evans, Jon Heron, Roshni R. Patel, and Nicola Wiles

Correspondence

Aripiprazole - data on efficacy and associated mortality
G. A. Silver, R. Gutierrez-Esteinou, and R. D. McQuade
Aripiprazole - data on efficacy and associated mortality
T. J. Crow

Author's reply
H. G. El-Sayeh, C. Morganti, and C. E. Adams

One hundred years ago

John Murray's Royal Asylum
Henry Rollin

Book reviews

Trauma, Truth and Reconciliation: Healing Damaged Relationships
Andrew Sims

Asylum to Action: Paddington Day Hospital, Therapeutic Communities and Beyond
Marianna Leontis

An Introduction to the Psychotherapies (4th edn)
Chris Mace

From the Editor's desk

From the Editor's desk
Peter Tyrer
The Late Dr Conolly, Resident Physician of Hanwell Lunatic Asylum (undated). Portrait by Messrs. Maull and Co. of Piccadilly and Cheapside. Picture selection and text by Dr Elizabeth Tovey, Central and North West London Mental Health Trust, and Dr Hagen Rampes, Barnet, Enfield and Haringey Mental Health NHS Trust.

John Conolly (1794–1866) was responsible for abolishing the use of restraint in the treatment of pauper patients at Hanwell Lunatic Asylum in the 1830s. He took the idea from Robert Gardiner Hill, a little-known house surgeon from Lincoln. John Conolly’s initial lack of success as both a clinician and a lecturer did not deter him and he secured a job as resident physician at Hanwell, a post that had previously eluded him. In his earlier work he advocated community care and used the term ‘restraint’ in its broader sense of both the use of mechanical constraints and the removal of patients from ordinary social life to confinement in an institution. He held that admission should only occur after a careful examination of the patient by a clinician with expertise in lunacy and that the asylum should be a place in which medical men were taught to recognise and treat mental disorder. Conolly’s achievements at Hanwell increased his reputation and he was eventually elected to the Fellowship of the Royal College of Physicians in 1844. In 1856 he wrote The Treatment of the Insane without Mechanical Restraints, which advocated ‘occupations’ in the daytime, ‘evening entertainments’ and treating patients with ‘kindness’. Hunter and Macalpine (Three Hundred Years of Psychiatry, 1963) judged Conolly one of the ‘outstanding figures’ in the history of psychiatry, although Andrew Scull (New Oxford Dictionary of National Biography, 2004) has provided a more skeptical view.

Thanks to Ealing Local History Centre, Ealing Central Library, 103 Ealing Broadway, London W5 5JY.
Highlights of this issue

BY SUKHWINDER S. SHERGILL

VIOLENCE: VICTIMS AND OFFENDERS

Asylum seekers have often been victims of violence. They are interviewed by the Home Office on arrival in the UK, and their disclosure of violence during this interview may be limited or absent. This is frequently revealed during subsequent interviews, when it can often be viewed sceptically. Data from Bogner and colleagues (pp. 75–81) suggest that guilt or shame may preclude victims of sexual violence from disclosing such personal information during these initial interviews. These victims also had higher levels of symptoms related to post-traumatic stress disorder. The authors suggest that late disclosure should not be assumed to reflect a fabricated asylum claim and there may be a need for more sensitive procedures during the initial assessment. An editorial by Herlihy & Turner (pp. 3–4) addresses the contemporary findings examining the relationship between memory and trauma, and how this may aid the assessment of victims of violence. The long-term outcome of patients discharged from medium secure care is relatively poor, with increased mortality, half of the patients being reconvicted and 38% sentenced to medium secure care is related to addiction. May play a central role in addiction. Williams et al. (pp. 3–4) address the contemporary findings examining the relationship between patients’ wishes and the services that may be available. They conclude that the challenges inherent in offering choice to service users should not preclude its application to psychiatry. One area where patient choice is especially relevant is in the field of addiction. Offering heroin rather than methadone has been suggested to be effective in severe dependence or where methadone maintenance has not been of benefit. Haasen et al. (pp. 55–62) demonstrate that heroin-assisted treatment was superior to methadone in terms of improved health and decreased illicit drug use. More heroin-assisted participants were retained in the study, although they suffered from a significantly increased rate of serious adverse effects. They suggest that heroin-assisted treatment may fulfil a useful role in treatment resistance. The neural basis of craving for drugs is of obvious interest in addictive disorders. Williams et al. (pp. 63–69) used positron emission tomography receptor imaging to demonstrate that there is an increased availability of opioid receptors in people with opioid dependence during early abstinence. Intriguingly, similar findings have been reported in alcohol and cocaine dependence, raising the issue that the opioid system may play a central role in addiction.

PATIENT CHOICE AND OPIOID DEPENDENCE

Increasing patient choice is expected to deliver better and more appropriate services. Samele and colleagues (pp. 1–2) describe the framework being developed for increased choice within mental health, focusing on supporting life choices, engagement with services, and choice in assessment and treatment. They highlight the difficulties applying these approaches to psychiatry, where the choices can be very limited for patients assessed as being at risk. Similarly, resource issues may also limit the fit between patients’ wishes and the services that may be available. They conclude that the challenges inherent in offering choice to service users should not preclude its application to psychiatry. One area where patient choice is especially relevant is in the field of addiction. Offering heroin rather than methadone has been suggested to be effective in severe dependence or where methadone maintenance has not been of benefit. Haasen et al. (pp. 55–62) demonstrate that heroin-assisted treatment was superior to methadone in terms of improved health and decreased illicit drug use. More heroin-assisted participants were retained in the study, although they suffered from a significantly increased rate of serious adverse effects. They suggest that heroin-assisted treatment may fulfil a useful role in treatment resistance. The neural basis of craving for drugs is of obvious interest in addictive disorders. Williams et al. (pp. 63–69) used positron emission tomography receptor imaging to demonstrate that there is an increased availability of opioid receptors in people with opioid dependence during early abstinence. Intriguingly, similar findings have been reported in alcohol and cocaine dependence, raising the issue that the opioid system may play a central role in addiction.

SELF-HARM AND DEPRESSION IN PREGNANCY

Against the background of an increasing rate of self-harm among young people, Young et al. (p. 44–49) found that young people outside the labour market, those who were unemployed, sick or not in full-time education, were the most likely to be engaged in chronic self-harm and actively trying to kill themselves. They suggest that targeting social causes may be more useful than biomedical intervention in this group. Depressive illness during pregnancy may have an impact on birth weight at term. A large longitudinal study by Evans et al. (pp. 84–85) did not support this hypothesis, showing that there was no significant association once the effects of health-related behaviours such as smoking had been adjusted for.
Patient choice in psychiatry

CHIARA SAMELE, SIMON LAWTON-SMITH, LESLEY WARNER and JEEVI MARIATHASAN

Summary The government has embarked on an ambitious plan to make patient choice central to the way healthcare and treatment are delivered. Mental healthcare is incorporated into this agenda. This editorial considers the implications of patient choice for psychiatry and some of the main challenges associated with this policy.

Declaration of interest None.

Choice of care is viewed as important to the modernisation of health and social care services, and has formed part of the government’s new delivery plan outlined in Creating a Patient Led NHS (Department of Health, 2005). Increasing choice is expected to create better alignment between what patients want and what services subsequently provide. It aims to promote greater patient autonomy, involvement and empowerment in the treatment and care received, to expand the range of available services, to help reduce waiting lists and to improve the quality of care through competition.

IMPLICATIONS OF CHOICE

The implications of patient choice are potentially huge for both patients and health and social care managers and staff. Choice places treatment or care decisions squarely with the patient. This is different to shared decision-making which involves at least two people (a clinician and a patient) agreeing which treatment option to implement (Charles et al, 1997). Informed choice is difficult to define and keep distinct from shared decision-making, participation or collaborative approaches. One definition includes ‘obtaining useful information from the practitioner or professional and then deciding individually or collaboratively on the best course of action that promotes independence, recovery and an improved quality of life’ (New York State Office of Mental Health, 2004). The provision of information alone, however, is not sufficient. It must be understood and presented in a balanced way so as not to suggest a right or wrong choice (Hope, 2002).

Critics of choice highlight concerns about the practical implementation and the potentially negative consequences to the patient. At an organisational level, creating the type of infrastructure required to support patient choice is complex. An effective health service based on choice requires fundamental changes to managerial and information systems, more time for consultations and a highly coordinated system to guide patients to appropriate care settings once choices have been made (Goodwin, 2006). At an individual level, Schwartz (2004) contends that too much choice can be debilitating, requiring more time to make decisions, with an increased risk of mistakes in decision-making and more negative psychological consequences to the patient.

CHOICE AND PSYCHIATRY

A framework has been developed which sets out the government’s vision for choice in mental health. This includes four ‘choice points’: promoting and supporting life choices (e.g. work, education, leisure, housing, self-help, direct payments); access and engagement (choice of how to contact mental health services, including in an emergency, and the role of advance directives); assessment (choice of when and where assessments take place); and informed choice of service or treatment and care pathway (including patients being supported to make their own decisions) (Care Services Improvement Partnership, 2006).

It might be particularly challenging for psychiatry to take on board this agenda for patient choice. To date acute physical healthcare and elective surgery are the main areas for patient choice. Initiatives such as ‘choose and book’ enable patients to select up to five different service providers and book appointments at preferred times. However, these initiatives might not be the best models for modern mental health services, whose ethos includes breaking down stigma and creating social inclusion by providing opportunities for employment and social activities (Valsra & Gardener, 2007). The recovery model for mental health underpins the choice agenda, in which a meaningful life can be lived despite a diagnosis of serious mental illness (Lester and Gask, 2006). Recovery seeks to work outside the medical model, and move away from a paternalistic approach to decision-making, to allow patients to regain independence and to access services that they feel best meet their needs.

A fundamental issue concerning patient choice within psychiatry is the dilemma posed by caring for patients and at the same time protecting them and society from harm. Of importance to psychiatrists is the patient’s capacity and competency to make valid treatment decisions. Using the example of anorexia nervosa, Henderson (2005) highlights how the capacity for choice and self-regulation of behaviour becomes a core part of treatment. He goes on to suggest that individuals are helped to regain their own volitional control, perhaps through cognitive psychotherapy. The danger, however, is that psychiatrists too readily assume that patients are not able to deal with information and choice. Hope (2002) suggests two methods to facilitate patient choice during a consultation: including patients’ values in the decision analysis and giving patients the necessary high-quality information to allow them to make informed decisions.

However, choices for those with mental illness can quickly become limited for those at high risk of harming themselves or others. For example, the application of the government’s proposed new powers of compulsory treatment, as set out in its Mental Health Bill 2006, will not take into account a patient’s capacity to make decisions about their medical treatment. It is unclear how compulsory community treatment in particular would coexist alongside patient choice, whether choice would act to reduce these powers or vice versa.
INTERNATIONAL LESSONS

What can we learn from how the choice agenda has been tackled elsewhere? Health departments from other high-income countries such as Australia, New Zealand, the USA and Canada broadly agree that patients should have more and better informed choice (Warner et al., 2006). In the USA it is accepted that consumer needs and choice should drive mental health services, but true choice is limited by the range of available services, and the complexity and lack of coordination between different agencies (statutory, voluntary and private).

In a list of ten ‘rules for quality mental health services in New York State’, rule number one states “There must be informed choice” (New York Office of Mental Health, 2005). Underpinning this document is a recovery-based principle in which informed choice includes obtaining useful information from the practitioner and an educational approach to medications and side-effects. However, a key problem identified is the limited willingness of many psychiatrists to collaborate about decisions concerning medication, citing their professional training or lack of capacity of the individual to make their own decisions as reasons.

In Australia, New Zealand and Canada a range of mental health plans, strategies and guidance refers to the importance of consumer participation (Warner et al., 2006). Key elements include adequate information for people to make informed choices, a range of alternative service providers and a recovery-based focus. However in practice choice is commonly not available. This might arise from health professionals’ reluctance to offer choices or through limitations on available services, primarily as a result of financial constraints both on services and on patients.

FUTURE DIRECTION

It is yet to be demonstrated whether patient choice will be fully embraced by psychiatry. The shift towards psychiatrists effectively handing over the reins to patients is likely to be gradual given the need to take account of issues such as capacity and risk. In addition, a better alignment between what patients want and what services they receive is dependent on factors, such as funding and service availability, which may be beyond psychiatrists’ control.

The profession would, however, leave itself open to fair criticism if it fails to engage with the government’s choice agenda. That agenda underpins much of the current reform in the National Health Service, and mental health patients should not be denied the possibility of the benefits that come from increased choice.

REFERENCES


Chiara Samele, PhD, Sainsbury Centre for Mental Health, London; Simon Lawton-Smith, BA, King’s Fund, London; Lesley Warner, RMN, RGN; Jeevi Mariathasan, PhD, Sainsbury Centre for Mental Health, London, UK

Correspondence: Dr Chiara Samele, Sainsbury Centre for Mental Health, 134–138 Borough High Street, London SE1 1LB, UK. Email: chiara.samele@schmh.org.uk

(First received 28 September 2006, final revision 30 January 2007, accepted 7 February 2007)


Asylum claims and memory of trauma: sharing our knowledge

JANE HERLIHY and STUART W. TURNER

Summary Decisions about asylum are extremely difficult because of the absence of a body of objective evidence. Psychiatrists and psychologists have a breadth of knowledge relating to the memory of trauma which could help to inform the asylum process, but we need to investigate how to apply this knowledge and how to make it accessible to decision makers.

Declaration of interest None.

In the UK, 25,710 applications for asylum were made in 2005 by people who claimed they were forced to leave their countries because of a fear of persecution. There were 33,940 appeals determined by the appeal court, the Asylum and Immigration Tribunal (Home Office, 2005). In order to make a claim, the asylum seeker usually has to relate a coherent account of events that they have experienced and which they claim led them to fear a return to their country. Of course, some will not have had the experiences that they allege – they will be presenting a false story in order to gain entry to the UK. The legal process for identifying valid claims involves written statements, interviews and court hearings, and is intended to identify those with a well-founded fear of persecution as defined in the 1951 Geneva Convention Relating to the Status of Refugees. One of the challenges with this jurisdiction is that decisions often have to be based on little more than the story that the claimant presents. The decision maker usually has to determine whether or not this is credible, in the absence of independent corroborating evidence about the applicant’s personal experience.

BREADTH OF EVIDENCE

The Working Party for Medical Evidence of the International Association of Refugee Law Judges (http://www.iarlj.nl) has as its remit to consider whether, and if so what, general principles apply to the form, reception and evaluation of expert evidence. Currently in the UK, expert reports are rarely requested by the authorities for asylum cases, and never by the courts; solicitors acting for the asylum seeker have discretion to request expert reports in individual cases. John Barnes, one of the founding members of the working party, has argued that, because medical evidence is usually written in the form of a report for one specific claimant, there is ‘no . . . breadth of evidence to assist in the evaluation of medical expert evidence’ (Barnes, 2004: p. 334). We propose, using the example of memory functioning, that there is a body of general evidence, but that more research is required into its application in this context.

MEMORY

One of the key factors when presenting a case for asylum is the ability of the asylum seeker to remember past experiences, usually traumatic, and give a coherent account of these to officials. A common assumption is that an experience of severe violence or torture will be so important that it will be remembered very clearly over the long term. If applicants for asylum change their account of their experiences (give discrepant accounts), this is therefore taken to suggest fabrication. This is an understandabe view but one which is challenged by scientific evidence.

Memory and trauma

When people witness accidents or crimes, some details are more likely to be remembered than others. Eyewitnesses to highly emotive events, such as violent crimes, tend to have a good memory for central details (moments central to the narrative or emotional gist of the event). However, these central details will be remembered at the expense of details peripheral to the overall unfolding of the event (Christianson & Safer, 1996).

We also know that when people are interviewed about what they have seen, it is possible to influence their answers. Questioning techniques used by the police and the courts have been modified following work on ‘suggestibility’ which shows that the wording of a question can influence the answer given, even by well-meaning witnesses (Gudjonsson, 1997).

When it comes to memories of personal experiences, we also know that emotion plays a big part both in what is encoded at the time and what is recalled later. The Yerkes–Dodson inverted-U model of performance and emotional arousal (Yerkes & Dodson, 1908; see Deffenbacher, 1983) reminds us that high levels of emotion may impair encoding of any memory, not just traumatic memories.

Many psychiatric disorders are known to be associated with aspects of memory. People who are depressed tend to have a memory bias for events which reflect negatively on themselves and the world, more easily forgetting situations in which they performed well. Anxiety is also associated with an attentional bias towards threatening situations or facts (Williams et al, 1997). Both depression and post-traumatic stress disorder (PTSD) have been shown to be associated with a pattern of overgeneral memory, in which individuals have difficulty retrieving memories of specific events (McNally et al, 1995; Williams, 1995).

Some memories of traumatic experiences are probably qualitatively different from normal autobiographical memories. An autobiographical memory for a normal event is verbal, sequenced (having a beginning, middle and end), recognised as being in the past, and may be recalled voluntarily. Traumatic memories often include incomplete autobiographical accounts. However, they often also include perceptual ‘snapshots’ (a smell, the sound of screaming, the image of a face), which are experienced in the present (reliving experiences) and are often triggered by external or internal cues (the sound of a firework, a feeling of guilt) rather than being subject to conscious recall (Hellawell & Brewin, 2004).

It is understandable that people faced with painful memories like these will adopt strategies to avoid situations likely to trigger them, for example meeting others from their country of origin. They may also adopt less conscious strategies such as ‘numbed’ emotions or dissociative amnesias.
Memory and the asylum process

Despite the long-established body of knowledge about memory, a review of the literature shows a dearth of articles by psychiatrists and psychologists for lawyers involved in decisions about those seeking asylum. Perhaps Barnes (2004) can be forgiven for his assertion about the lack of a breadth of evidence.

We searched PsydINFO, Medline and PILOTS (the database of the National Center for Post-Traumatic Stress Disorder; http://www.ncptsd.va.gov/ncmain/index.jsp) with the terms memory AND trauma AND law AND (refugees OR asylum). Only three papers were identified that explicitly linked memory functioning to asylum decision-making.

Masinda (2004) analysed a series of seven negative asylum decisions on refugees exhibiting PTSD, comparing judicial determinations with clinical and research findings on memory. Herlihy et al (2002) interviewed on two occasions refugees granted asylum as a group by the United Nations and found inconsistency between their accounts on the two occasions. They found a relationship between the rate of discrepancies and the nature of the questions asked. Furthermore, individuals with higher levels of PTSD were more inconsistent the longer they had to wait between interviews. Morgan et al (2004) studied over 500 soldiers undergoing ‘high-stress’ interrogation, ‘modelled from the experience of actual military personnel who have been prisoners of war’. These young, fit, trained individuals managed only a 66% recognition rate when presented with photographs (in identical clothes to improve performance) of their interrogators.

Other issues are probably also important in this context. For example, from clinical experience and the research literature, we know that when people feel shame they find it difficult to disclose personal information. In an interview with the authorities, however, there are often opposing forces at work. Shame regarding a brutal rape, for example, is likely to inhibit disclosure of the event to an official (Van Velsen et al, 1996), yet disclosure may be essential to gain protection from the possibility of further brutality.

Claimants’ discrepant accounts of their experiences may also be related to the different triggering of traumatic memories depending on the situation. We do not know enough about the impact of this on the variability of presentation in the highemotion contexts of the court room, the official interview and clinical assessment.

Some of these questions are being addressed. Steel et al (2004) used case examples to illustrate their arguments that mental health issues have an impact on the fairness of refugee status decisions. A recent study explored the experiences of 27 asylum seekers of interviews at the UK Home Office and identified the importance of the behaviour of interviewers as well as the impact of claimants’ feelings of shame and efforts to avoid their memories (Bogner et al, 2007, this issue).

FUTURE DIRECTIONS

Psychiatrists and psychologists practising in this field hold a wealth of clinical knowledge which may be relevant to the legal process of deciding asylum claims. Legal advisors and immigration judges sometimes look to medical experts to help them to make these very difficult decisions. We need to find ways of developing the broader evidence base concerning not only memory and the asylum process, but also the impact of traumatic experiences, cross-cultural assessments, depression, stressed environments and suicide risk assessments. We then need to make this evidence more accessible to decision makers. If we achieve this, we will have helped to produce a more robust system (with fewer false positives as well as false negatives), one better able to achieve fair decisions for all.

REFERENCES


Theory of mind in schizophrenia

Meta-analysis

MIRJAM SPRONG, PATRICIA SCHOTHRST, ELLEN VOS, JOOP HOX and HERMAN VAN ENGELAND

Background Mentalising impairment (an impaired ability to think about people in terms of their mental states) has frequently been associated with schizophrenia.

Aims To assess the magnitude of the deficit and analyse associated factors.

Method Twenty-nine studies of mentalising in schizophrenia (combined n=1518), published between January 1993 and May 2006, were included to estimate overall effect size. Study descriptors predicted to influence effect size were analysed using weighted regression-analysis techniques. Separate analyses were performed for symptom subgroups and task types.

Results The estimated overall effect size was large and statistically significant (d = -1.255, P < 0.0001) and was not significantly affected by sample characteristics. All symptom subgroups showed significant mentalising impairment, but participants with symptoms of disorganisation were significantly more impaired than the other subgroups (P < 0.01).

Conclusions This meta-analysis showed significant and stable mentalising impairment in schizophrenia. The finding that patients in remission are also impaired favours the notion that mentalising impairment represents a possible trait marker of schizophrenia.

Declaration of interest None. Funding detailed in Acknowledgements

‘Theory of mind’ and ‘mentalising’ refer to the cognitive ability to attribute mental states such as thoughts, beliefs and intentions to people, allowing an individual to explain, manipulate and predict behaviour. In 1992 Frith proposed a relationship between theory of mind and schizophrenia, and argued that several symptoms of schizophrenia could be explained by mentalising impairment (Frith, 1992). This led to a substantial body of research which has recently been critically reviewed twice (Brune, 2003a; Harrington et al, 2005a).

In both reviews it was concluded that theory of mind is impaired in individuals with schizophrenia. Although these reviews were executed thoroughly, they are limited to a qualitative description of the observed deficit, thus lacking important information on the magnitude of the effect. The purpose of this meta-analysis is to produce a synthesised effect size estimate that has considerably more power than the individual studies. In addition, effects of study characteristics on the findings are analysed.

METHOD

Study selection
An extensive literature search was conducted in the electronic databases Medline, EMBASE and PsycINFO (January 1993 to May 2006) using the following key words: theory of mind, mentalizing, social cognition, schizophrenia, and psychosis. Additional studies were identified by checking the reference lists from identified reviews and papers on the topic. To ensure that we did not overlook studies published by May 2006 but not included in the computerised databases by that date, a journal-by-journal search was performed in the January 2006 to May 2006 editions of the American Journal of Psychiatry, Biological Psychiatry, Journal of Nervous and Mental Disease, Psychiatry Research, Schizophrenia Bulletin, Schizophrenia Research and Psychological Medicine. Studies considered eligible for this meta-analysis were empirical research studies written in the English language and published in peer-reviewed journals. Research samples had to be composed of adults diagnosed with schizophrenia or schizoaffective disorder according to the established diagnostic systems (DSM or ICD). Their sample group’s mentalising performance had to be compared with that of healthy controls. Measures of mentalising included in this meta-analysis are described below. Finally, sufficient data had to be reported for the computation of the standardised mean difference (Lipsey & Wilson, 2001).

Types of mentalising tasks
There is a fair amount of agreement on the definition of theory of mind among researchers. However, this definition is broad, perhaps reflecting the fact that it is probably not a unitary function. This has led to a wide variation in the operationalisation of the concept. One of the most frequently used types of mentalising tasks is the false belief or deception task (e.g. Frith & Corcoran, 1996; Corcoran et al, 1997; Doody et al, 1998; Mazza et al, 2001). In a first-order false belief/deception task, the ability to understand that someone can hold a belief that is different from the actual state of affairs is assessed. In a second-order false belief/deception task, participants have to infer the (false) beliefs of one character about the (false) beliefs of a second character.

A second type of theory of mind task commonly used in schizophrenia research is an intention-inferencing task, in which the ability to infer a character’s intentions from information in a short story is assessed (e.g. Sarfati et al, 1997a,b, 1999, 2000; Sarfati & Hardy-Baylé, 1999). A third type of task measures the ability to understand indirect speech, such as in irony, banter, hints and metaphors (e.g. Corcoran et al, 1995; Langdon et al, 2002; Corcoran, 2003; Corcoran & Frith, 2003; Craig et al, 2004). This is based on the notion that for the understanding of indirect speech an understanding of another person’s mental state is required (e.g. Sperber & Wilson, 2002). However, Langdon & Coltheart (2004) showed that comprehension of irony and comprehension of metaphors are unrelated and that having an intact theory of mind is a prerequisite for the interpretation of irony but not for...
the interpretation of metaphors. Therefore, data on the interpretation of metaphors were excluded from this meta-analysis.

A fourth, less commonly used type of theory of mind task in schizophrenia research is the attribution of mental states to animated geometric shapes which interact in a ‘socially’ complex way (Blakemore et al., 2003; Russell et al., 2006). This type of task may not be fully comparable with the other theory of mind tasks because of the higher level of abstraction involved. Finally, in some studies the ‘eyes’ task is used, in which participants have to infer mental states from looking at pictures of eyes (Kington et al., 2000; Russell et al., 2000; Kelemen et al., 2005). This has been referred to as a theory of mind task, but at face value the construct being measured seems to be different from that assessed by the other paradigms, perhaps assessing emotion recognition abilities or empathy rather than theory of mind.

Since there is a serious lack of research on the psychometric properties (including construct validity and criterion validity) of the many different theory of mind tasks that have been developed (Harrington et al., 2005a), it may not be possible to formulate completely objective inclusion criteria regarding the type of tasks used in the studies. In this meta-analysis this problem is addressed statistically in two ways. First, homogeneity analyses are used to check whether the grouping of effect sizes from different studies shows more variation than would be expected from sampling error alone, indicating that the effect sizes may not be comparable. A second approach to this problem is to break down the overall mean effect size into mean effect sizes for different types of tasks. For these mean effect sizes per type of task to be meaningful, we (subjectively) set a minimum of five eligible studies per sub-task analysis. This led to the exclusion of two studies using tasks assessing the attribution of mental states to abstract shapes rather than humans (Blakemore et al., 2003; Russell et al., 2006), and three studies in which the ‘eyes’ task was used (Kington et al., 2000; Russell et al., 2000; Kelemen et al., 2005).

**Schizophrenia subgrouping**

Ever since Frith’s first proposal (Frith, 1992), the association between mentalising and the core symptoms of schizophrenia has been an important focus of research interest. Schizophrenia is a heterogeneous disorder and various subgrouping methods have been used, based on different theories regarding the relationship between mentalising and symptomatology.

In earlier studies, Frith and colleagues divided their schizophrenia samples into six symptom subgroups (Corcoran et al., 1995). In their later studies, the number of subgroups was reduced to four, categorised as follows:

(a) behavioural signs of negative symptoms and/or incoherence;

(b) paranoid symptoms (delusions of persecution, delusions of reference, and third-person hallucinations);

(c) passivity experiences (delusions of control, thought insertion, and thought broadcasting);

(d) symptoms in remission.

The first group was predicted to be the most impaired, because of these patients’ incapacity to represent the mental states of others as well as themselves. Paranoid patients would perform poorly because of their difficulties in monitoring other people’s intentions. Patients whose symptoms were in remission and patients with passivity symptoms were predicted to have normal mentalizing abilities. These hypotheses were largely confirmed and have repeatedly been replicated (Frith & Corcoran, 1996; Corcoran et al., 1997; Pickup & Frith, 2001).

Sarfati and colleagues (Sarfati et al., 1997a,b, 1999; Sarfati & Hardy-Bayle, 1999) and Zalla et al (2006) suggested that impairment of theory of mind is related to thought disorder, reflecting an executive functioning deficit. Thus, their samples were divided into those with and those without thought disorder. In all of their studies thought-disordered participants performed significantly more poorly than healthy controls. However, in two of the studies the non-disorganised participants also showed poor performance (Sarfati et al., 1997b; Zalla et al., 2006).

Three research groups studied the relationship between mentalising and paranoid delusions (Randall et al., 2003; Craig et al., 2004; Harrington et al., 2005b). In all three studies patients with paranoid delusions showed impairment of theory of mind relative to the normal control group. However, in the study by Randall et al (2003), theory of mind performances of the paranoid and non-paranoid subgroups did not differ significantly from each other.

Lastly, Herold et al (2002) investigated whether the deficit in theory of mind was state- or trait-dependent and therefore assessed patients whose schizophrenia was in remission. Results showed that theory of mind impairment was still present in the remission phase of the illness.

**Moderator variables**

Published research suggests a number of variables that may affect mentalising performance and thus influence effect size. Hence, we aimed to code these variables in order to evaluate their influence on the effect size. Potential moderator variables at individual patient level are age, gender, medication, IQ, disease status (acute, chronic or in remission), severity of psychopathology, and symptoms. To analyse the effect of specific clusters of symptoms on mentalising impairment, the symptom subgroups used by different research groups were divided into four categories:

(a) symptoms of disorganisation;

(b) no symptoms of disorganisation;

(c) paranoid symptoms;

(d) remitted patients.

The disorganised subgroup was composed of the behavioural symptoms subgroup of the studies by Frith and colleagues (Corcoran et al., 1995, 1997; Pickup & Frith, 2001) and the disorganised subgroups of the Sarfati, Mazza and Zalla studies (Sarfati et al., 1997a,b, 1999; Sarfati & Hardy-Bayle, 1999; Mazza et al., 2001; Zalla et al., 2006). The non-disorganised patients of the Sarfati and Zalla studies were combined into the second subgroup (Sarfati et al., 1997a,b, 1999; Sarfati & Hardy-Bayle, 1999; Zalla et al., 2006). For the paranoid subgroup the results of the studies focusing on paranoid schizophrenia (Randall et al., 2003; Craig et al., 2004; Harrington et al., 2005b) were combined with the results for the paranoid subgroups of the studies by Frith and colleagues (Corcoran et al., 1995, 1997; Pickup & Frith, 2001). The remitted disease subgroup comprised the patients in remission in the studies by Herold et al (2002), Randall et al (2003) and Frith and colleagues (Corcoran et al., 1995; Corcoran et al., 1997; Pickup & Frith, 2001). The passivity subgroup of Frith and colleagues was not coded, because results for that subgroup were reported only in two studies.

Potential moderators at study level are the matching of patients and controls on
group characteristics (e.g. mean age, mean IQ, gender distribution), type of mentalising task used, and whether the task is administered verbally or non-verbally. Four types of theory of mind tasks were distinguished: first-order false belief/deception; second-order false belief/deception; intention inferencing; and comprehension of indirect speech. Some tasks did not fit in any of these categories, for example the false belief/deception tasks for which the orders were unknown or mixed.

Within the different task paradigms there is also variation in whether tasks are presented in a verbal or non-verbal form. It has been suggested that verbalisation may be impoverished in schizophrenia and may constitute an experimental bias in favour of a theory of mind deficit in people with schizophrenia (e.g. Sarfati et al, 1999). In a separate coding, tasks were classified as verbal or non-verbal.

Coding
Each study was coded independently by two authors (M.S. and E.V.). In case of discrepancies, consensus was reached in conference with the whole research group. When results were reported in graphical form only an email was sent to the author with a request for the exact numerical results.

Data collection and analysis
For each study an unbiased standardised mean difference \(d\) was calculated using reported means and standard deviations. This effect size statistic is computed as the difference between the mean of the schizophrenia group and the mean of the control group, divided by the pooled standard deviation. Hedges' formula was applied to correct for upwardly biased estimation of the effect size in small samples (Lipsey & Wilson, 2001).

When means and standard deviations were not available, \(d\) was calculated from the reported \(t\) or \(F\) values. In cases where the only reported outcome variable was the proportion of participants with a good (or poor) performance, \(d\) was estimated using the probit transformation method (Lipsey & Wilson, 2001). A sensitivity analysis was performed to check whether there was any significant effect of using probit-transformed effect sizes on the overall effect size. In studies in which data were reported for (symptom) subgroups only, data were first pooled and then compared as one group with the control group. In addition, the effect sizes of symptom subgroups were calculated for subsequent analyses. Several studies used more than one (sub)task to assess theory of mind, and therefore had more than one effect size; in these cases a pooled effect size was computed. However, if the authors had included a composite score, the effect size of this score was calculated. Again, effect sizes for different task types were calculated for subsequent analyses. In addition to the individual effect sizes and 95% confidence intervals, \(P\) values were calculated for each study using two-tailed independent \(t\)-tests and \(\chi^2\)-tests.

The mean effect size across studies was calculated by weighting each effect size by the inverse of its sampling variance. A confidence interval and \(z\)-value were calculated to examine the statistical significance of the effect. To test whether the individual effect sizes are good estimators of the population effect size, the homogeneity statistic \(Q\) was calculated (Lipsey & Wilson, 2001). Because sample sizes are small in the subgroup and task type analyses (see below), a random effects model was fitted to the data (Lipsey & Wilson, 2001). To examine publication bias, a fail-safe number was computed using Orwin's formula (Lipsey & Wilson, 2001). This indicates the number of studies with null effects that have to reside in file drawers to reduce the mean effect size to a negligible level (which we set at 0.2). Weighted regression analysis was performed using the statistical package Meta-Stat (Rudner et al, 2002) to evaluate whether group differences in IQ, gender and age had an impact on effect size. Other variables with a potential influence on effect size, such as patient status, medication use and severity of psychopathology, could not be analysed because of the small number of studies reporting results for these parameters. Separate analyses were performed to analyse whether mentalising impairment is different for different symptom subgroups or for different types of mentalising tasks.

RESULTS
The literature search resulted in a total of 32 studies meeting the inclusion criteria. One publication (Langdon et al, 2002a) was excluded because data concerning the same participants had been reported in another paper (Langdon et al, 2002b).

Sample characteristics (\(n\), mean age, percentage of males, mean score on the Binois–Pichot Vocabulary Scale and mean score on the non-verbal theory of mind test) were exactly the same in two studies by Sarfati and colleagues (Sarfati & Hardy-Bayle, 1999; Sarfati et al, 2000), suggesting that the same patient samples had been used. Because in the first of these studies the patient sample was divided into symptom subgroups, but more control participants and an additional theory of mind task were used in the latter study, instead of selecting one of the two studies the results of both were combined. Because we were unable to contact the authors of one study within the time frame of data collection and data analysis to obtain the exact numerical results which were not reported in the article, the results of that study could not be included in the meta-analysis (Frith & Corcoran, 1996). The characteristics of the remaining 29 studies with a total of 831 patients (mean age 35.9 years, 70% male, mean IQ 98.7) and 687 controls (mean age 35.2 years, 60% male, mean IQ 105.3) are listed in Table 1.

Analysis of the total sample
Figure 1 shows the 29 individual effect sizes with their 95% confidence intervals. None of the confidence intervals includes the value zero, indicating a statistically significant effect for each study. The weighted mean effect size of the combined sample is \(-1.255\) (95% CI \(-1.441\) to \(-1.069\)) which is also statistically significant \((z=13.25, P<0.0001)\). Homogeneity analysis showed that there was homogeneity among studies \((Q=29.13, \text{ d.f.}=28, P=0.41)\), and weighted regression analysis did not show any relationship between effect size and difference between patient and control groups in IQ \((P=0.193)\), proportion of males \((P=0.115)\) and age \((P=0.147)\). The fail-safe number was 153, which indicates that 153 unpublished studies are required to reduce the effect size of the combined findings to a negligible level.

Analyses of the symptom subgroups
Mean effect sizes and confidence intervals of the symptom subgroups are displayed in Fig. 2. The disorganised patients performed worst on the mentalising tasks compared with healthy controls \((d=-2.231, 95\% \text{ CI } -2.565\) to \(-1.897,\)
Table 1  Summary of studies included in the meta-analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Schizophrenia/control sample</th>
<th>Schizophrenia subgroups</th>
<th>Mentalising tasks</th>
<th>p^2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean age (years)</td>
<td>Males (%)</td>
<td>Mean IQ</td>
</tr>
<tr>
<td>Corcoran et al (1995)</td>
<td>55/30</td>
<td>32/31</td>
<td>69/67</td>
<td>98/107</td>
</tr>
<tr>
<td>Sarfati et al (1997a)</td>
<td>24/24</td>
<td>32/32</td>
<td>79/58</td>
<td>NR</td>
</tr>
<tr>
<td>Sarfati et al (1997b)</td>
<td>12/12</td>
<td>27/26</td>
<td>42/50</td>
<td>NR</td>
</tr>
<tr>
<td>Corcoran et al (1997)</td>
<td>44/40</td>
<td>30/32</td>
<td>71/43</td>
<td>102/108</td>
</tr>
<tr>
<td>Langdon et al (1997)</td>
<td>20/20</td>
<td>33/NR</td>
<td>45/NR</td>
<td>NR</td>
</tr>
<tr>
<td>Doody et al (1998)</td>
<td>28/20</td>
<td>46/20</td>
<td>61/45</td>
<td>108/109</td>
</tr>
<tr>
<td>Sarfati (2000)</td>
<td>25/25</td>
<td>33/NR</td>
<td>28/NR</td>
<td>NR</td>
</tr>
<tr>
<td>Mazza et al (2000)</td>
<td>55/17</td>
<td>34/37</td>
<td>89/86</td>
<td>88/90</td>
</tr>
<tr>
<td>Pickup &amp; Frith (2001)</td>
<td>40/35</td>
<td>39/43</td>
<td>73/54</td>
<td>93/103</td>
</tr>
<tr>
<td>Langdon &amp; Coltheart (2001)</td>
<td>32/24</td>
<td>37/35</td>
<td>56/50</td>
<td>NR</td>
</tr>
<tr>
<td>Langdon et al (2002b)</td>
<td>25/20</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Herold et al (2002)</td>
<td>20/20</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Janssen et al (2003)</td>
<td>43/43</td>
<td>32/35</td>
<td>56/51</td>
<td>105/113</td>
</tr>
<tr>
<td>Corcoran &amp; Frith (2003)</td>
<td>59/44</td>
<td>41/40</td>
<td>85/80</td>
<td>101/103</td>
</tr>
<tr>
<td>Brunet et al (2003a)</td>
<td>7/8</td>
<td>31/23</td>
<td>100/100</td>
<td>111/119</td>
</tr>
<tr>
<td>Brunet et al (2003b)</td>
<td>25/25</td>
<td>31/34</td>
<td>76/68</td>
<td>NR</td>
</tr>
<tr>
<td>Mazza et al (2003)</td>
<td>39/20</td>
<td>43/43</td>
<td>83/65</td>
<td>87/86</td>
</tr>
<tr>
<td>Brune (2003)</td>
<td>23/12</td>
<td>29/30</td>
<td>74/58</td>
<td>92/107</td>
</tr>
<tr>
<td>Corcoran (2003)</td>
<td>39/44</td>
<td>41/40</td>
<td>82/80</td>
<td>103/104</td>
</tr>
<tr>
<td>Randall et al (2003)</td>
<td>32/18</td>
<td>35/32</td>
<td>69/61</td>
<td>110/115</td>
</tr>
<tr>
<td>Craig et al (2004)</td>
<td>16/16</td>
<td>32/29</td>
<td>69/69</td>
<td>105/110</td>
</tr>
<tr>
<td>Marjoram et al (2005a)</td>
<td>20/20</td>
<td>40/40</td>
<td>60/55</td>
<td>97/100</td>
</tr>
<tr>
<td>Marjoram et al (2005b)</td>
<td>15/15</td>
<td>28/34</td>
<td>87/67</td>
<td>97/106</td>
</tr>
</tbody>
</table>

(Continued)
The confidence interval of the mean effect size in the disorganized subgroup shows no overlap with that in the non-disorganized (d = 1.278, 95% CI 1.771 to 0.785, P < 0.01) and paranoid subgroups (d = 1.241, 95% CI 1.514 to −0.968, P < 0.01), indicating that the difference between the disorganized subgroup and the other subgroups is statistically significant. This was confirmed by post hoc comparisons of the mean effect size of the disorganized subgroup v. the mean effect sizes of the other three symptom subgroups (all P values < 0.01). Interestingly, patients in remission also showed a significantly worse performance than controls (d = −0.692, 95% CI −1.017 to −0.367, P < 0.01). The homogeneity statistic of the non-disorganized subgroup was statistically significant (Q = 7.3816, d.f. = 4, P < 0.05), indicating that the effect sizes within this

patients in remission also showed a significantly worse performance than controls (d = −0.692, 95% CI −1.017 to −0.367, P < 0.01). The homogeneity statistic of the non-disorganized subgroup was statistically significant (Q = 7.3816, d.f. = 4, P < 0.05), indicating that the effect sizes within this
subgroup analysis differed more than would be expected from sampling error alone, perhaps owing to differences associated with study (or sample) characteristics. This was somewhat surprising, since four of the five studies were by the same research group. The finding that the other three homogeneity statistics were not statistically significant suggests that although different authors might have used different criteria for their symptom subgroups, combining these subgroups was meaningful.

Analyses of the types of mentalising tasks

The mean effect sizes and confidence intervals of the four theory of mind task categories are shown in Fig. 3. The mean effect sizes of the first-order tasks \( d = -1.193, 95\% \text{ CI} -1.666 \text{ to } -0.720, P < 0.01 \) and the second-order tasks \( d = -1.443, 95\% \text{ CI} -1.867 \text{ to } -1.019, P < 0.01 \) have homogeneity statistics indicating heterogeneity among the effect sizes: \( Q = 97.691, d.f. = 12 (P < 0.01) \) and \( Q = 17.875, d.f. = 6 (P < 0.01) \) respectively. In contrast, the mean effect sizes within both the indirect speech tasks \( d = -1.040, 95\% \text{ CI} -1.301 \text{ to } -0.779, P < 0.01 \) and the intention-inferencing tasks \( d = -0.959, 95\% \text{ CI} -1.228 \text{ to } -0.690, P < 0.01 \) are both homogeneous. The difference between the mean effect sizes for different subtasks could not be analysed statistically, because not all effect sizes were statistically independent since in one study different types of tasks might have been used.

The mean effect size of studies using verbal tasks is comparable with the mean effect size of studies using non-verbal tasks (verbal, \( d = -1.221, 95\% \text{ CI} -1.462 \text{ to } -0.980 \); non-verbal \( d = -1.251, 95\% \text{ CI} -1.496 \text{ to } -1.006 \)). The homogeneity statistics of the verbal and non-verbal tasks both show heterogeneity among the effect sizes. Again, the difference could not be analysed because of statistical dependence.

**DISCUSSION**

The aim of this meta-analysis was to investigate the extent of mentalising impairment in people with schizophrenia. By combining 29 studies, a total sample size was created of over 1300 participants. The overall effect size was \(-1.1255\), indicating that on average the theory of mind performance of participants with schizophrenia is more than one standard deviation below that of healthy controls. According to a widely used convention for appraising the magnitude of effect sizes this is considered a large effect (Cohen, 1988). Homogeneity analysis showed that the mean effect size of the combined samples is a good estimate of the typical effect size in the population. The large fail-safe number makes the ‘file drawer’ problem, which is a limitation of some meta-analyses, negligible.

The moderator variables IQ, gender and age did not significantly affect mean effect size. Thus, the impairment in theory of mind is robust and is not readily moderated by variables that may seem relevant. However, the effect of other potentially important moderator variables such as medication use and duration and severity of illness could not be analysed owing to a lack of information on these characteristics in many studies.

Participants with schizophrenia who had signs and symptoms of disorganisation were found to be significantly more impaired in terms of theory of mind than those in the other symptom subgroups. However, these results may also be explained by the composition of the disorganised symptom subgroup. The behavioural subgroup of the studies by Frith and colleagues was ranked highest in their hierarchical model. Thus, individuals in this group might also have had symptoms of the paranoid and/or passivity subgroup. This brings the risk that poorer performance in this group may be explained by having more severe and complex symptoms (Harrington et al, 2005a). Similarly, in two of the four studies by Sarfati and colleagues the disorganised subgroup had more general psychopathology, which might explain their poorer theory of mind performance (Sarfati & Hardy-Bayle, 1999; Sarfati et al, 1999).

The mean effect size \( d = -0.692 \) of mentalising impairment in patients in remission was smaller than in the other symptom subgroups, but is still considered to be medium to large (Cohen, 1988). Moreover, this effect did not differ significantly from the effect sizes of the disorganised and paranoid subgroups.

Unexpectedly and despite apparent differences in type and difficulty of the theory of mind tasks – the mean effect sizes for different task types were found to be similar. An explanation might be that our method of grouping studies by task types was not correct. This is supported by the finding that two of the four task type analyses showed heterogeneity among effect sizes. However, since there is a lack of research on the psychometric properties of the tasks that were used, such as construct and concurrent validity, it is not yet possible to categorise these tasks objectively.

There was also no difference between the mean effect sizes of verbal and non-verbal tasks, which is consistent with the findings of Sarfati and colleagues (Sarfati et al, 1999, 2000). Thus, impairment of theory of mind does not appear to be
affected by verbalisation deficits that have been reported in people with schizophrenia.

**Mentalising in schizophrenia: generalised v. specific impairment**

As shown by Heinrichs & Zakzanis (1998), people with schizophrenia show generalised neurocognitive impairment. On their list of 22 mean effect sizes of common neurocognitive tests, the effect size of mentalising impairment would be ranked fourth. An interesting question is whether poor mentalising performance in schizophrenia interacts with or is influenced by general cognitive impairment. This problem is acknowledged by some authors, who corrected for general cognitive abilities by matching groups on IQ, covarying out cognitive variables (e.g. attention, executive functioning, memory, general picture sequencing abilities) or excluding participants from statistical analyses if they answered reality questions about the theory of mind stories incorrectly. In their reviews, Brune (2005a: p. 25, Table 1) and Harrington et al. (2005a: pp. 252–267, Table 1) discussed the empirical evidence as to whether the mentalising deficits in schizophrenia are specific or the consequence of general cognitive impairment. In both reviews it was concluded that the evidence speaks in favour of the notion that there is a specific theory of mind deficit in schizophrenia. As with many neurocognitive tests, theory of mind tasks probably measure several component processes at the same time. For example, tasks in which the comprehension of indirect speech is assessed may require not only mentalising abilities but also basic language comprehension and expressive language skills. Possibly, general cognitive abilities represent a necessary but not sufficient condition for adequate mentalising, which is known as the ‘building block’ view of social cognition (see Penn et al., 1997).

**Mentalising in schizophrenia: state or trait dependency**

In his cognitive model of the relationship between meta-representation and the signs and symptoms of schizophrenia, Frith assumed that in people with this disorder, the initial development of mentalising abilities is relatively normal and that these abilities become impaired as the illness develops (Frith, 1992). In the subsequent studies by him and his colleagues, it was predicted and found that patients who were in remission (i.e. symptom-free) were unimpaired compared with normal controls (e.g. Corcoran et al., 1995, 1997; Frith & Corcoran, 1996; Pickup & Frith, 2001). In contrast, our meta-analysis has shown that patients have significant impairment during remission, which is consistent with the findings of Herold et al. (2002). These findings support the notion that mentalising is not just a consequence of the acute phase of the disorder but may be trait-dependent. It cannot be excluded that the criteria for remission (e.g. partial or full remission) used by Herold et al. (2002) and by Frith and colleagues are different. Other factors such as (prophylactic) treatment may also explain the divergent findings. However, more support for the trait argument comes from studies on mentalising in populations at elevated risk of developing a psychotic illness.

In general, people at genetic risk of schizophrenia show reduced performance on the more common types of theory of mind tasks (Wykes et al., 2001; Irani et al., 2006; Marjoram et al., 2006), but not on the ‘eyes’ test (Kelemen et al., 2004; Irani et al., 2006). In the study by Schiffman et al. (2004), genetic high-risk children who would later develop schizophrenia-spectrum disorders had lower scores on a role-taking task, which the authors considered assessed a facet of theory of mind. An association between theory of mind performance and subclinical schizotypal traits has also been found (Langdon & Coltheart, 1999, 2004; Irani et al., 2006; Meyer & Shean, 2006). Pickup (2006) showed that schizotypal traits analogous to positive symptoms of schizophrenia predicted poorer mentalising performance, whereas no association was found between poorer theory of mind and schizotypal traits analogous to the ‘behavioural signs’ of schizophrenia. Platek et al. (2003) suggested that contagious yawning is part of a more general phenomenon known as mental state attribution. Consistent with this hypothesis, susceptibility to contagious yawning was positively related to performance on (other) mentalising tasks, and negatively related to schizotypal personality traits. Only in the study by Jahshan & Sergi (2007) was there no difference between people with high schizotypy and those with low schizotypy regarding theory of mind performance. There is thus considerable evidence that mentalising impairment is a susceptibility indicator for schizophrenia and hence may be trait-dependent.

**Limitations**

The first limitation, to which we have already alluded, is that studies were excluded in which less common types of theory of mind tasks were used. Because there is no information on the psychometric properties of the many different tasks, this is somewhat arbitrary. In addition, the categorisation of task type is not supported by psychometric evidence. Second, the method of categorising symptom subgroups employed in this meta-analysis should be considered tentative. The main problem with our approach is that there is overlap between symptom clusters; for example, the subgrouping method used by Frith and colleagues is hierarchical, with the behavioural subgroup being the highest category. This means that patients in that subgroup could also report paranoid symptoms, but those in the paranoid subgroup could not report behavioural symptoms. As another example, participants categorised as paranoid in the study by Harrington et al. (2005b) could also have formal thought disorder (which was indeed the case). However, in spite of this limitation, we believe that the results of the subgroup analyses in this meta-analysis are valuable. This is statistically supported by the homogeneity analyses, which show that the clustering of symptom subgroups did not result in more variation than would be expected from sampling error alone and that it is plausible that the studies within the subgroup analyses are comparable.

**Recommendations for future research**

The results and limitations of this meta-analysis lead to some recommendations for future research. First, research focusing on the mentalising process itself is necessary, addressing questions on what components it comprises and on how to operationalise them. As has already been pointed out by Harrington et al. (2005a), it is also important to establish the psychometric properties of theory of mind tasks. Second, the finding that the deficit in theory of mind in schizophrenia is perhaps trait-dependent rather than state-dependent implies that the deficit may also be present before illness onset. Therefore, there may be a role of mentalising impairment in the early detection and prediction of schizophrenia, requiring a longitudinal study examining theory of mind abilities in people at risk of developing schizophrenia.
The finding that theory of mind impairment may be trait-dependent also brings to mind a comparison with autism-spectrum disorders. An impaired ability to understand mental states has been described as one of the core symptoms of such disorders (Yirmiya et al., 1998). However, although the risk of psychotic disorder is elevated in individuals with autism-spectrum disorder (Stahlberg et al., 2004), most of them will not develop a psychotic disorder. Future research should focus on what regard to theory of mind in these disorders. Abu-Akel & Bailey (2000) for example suggested that there might be different forms of impairment of theory of mind. They argue that, unlike people with autism-spectrum disorders, people with schizophrenia do not lack an understanding that others have mental states; instead, they may overattribute knowledge to others or apply their knowledge of mental states in an incorrect or biased way. Thus, an interesting research topic would be a comparison of the mentalising abilities of groups of people with these two disorders.

Lastly, social impairment is one of the most disabling clinical features of schizophrenia and it is well known that it is often present before illness onset (e.g. Niemi et al., 2003). Since theory of mind impairment appears to be trait- rather than state-dependent in schizophrenia, this deficit may have a role in the development of social impairment. However, evidence of a relationship between theory of mind performance and social functioning is lacking and should be an aim of future research.

ACKNOWLEDGEMENTS

This study was supported by a grant from ZorgOnderzoek Nederland/NWO-Medische Wetenschappen, project 2630.0001.

REFERENCES


Brunet, E., Sarfati, Y. & Hardy-Bayle, M. C. (2003b) Reasoning about physical causality and other’s intentions in schizophrenia. Cognitive Neuropsychiatry, 8, 129–139.


Cost-effectiveness of first- v. second-generation antipsychotic drugs: results from a randomised controlled trial in schizophrenia responding poorly to previous therapy

L. M. DAVIES, S. LEWIS, P. B. JONES, T. R. E. BARNES, F. GAUGHRAN, K. HAYHURST, A. MARKWICK and H. LLOYD on behalf of the CUTF.LASS team

Background There are claims that the extra costs of atypical (second-generation) antipsychotic drugs over conventional (first-generation) drugs are offset by improved health-related quality of life.

Aims To determine the relative costs and value of treatment with conventional or atypical antipsychotics in people with schizophrenia.

Method Cost-effectiveness acceptability analysis integrated clinical and economic randomised controlled trial data of conventional and atypical antipsychotics in routine practice.

Results Conventional antipsychotics had lower costs and higher quality-adjusted life-years (QALYs) than atypical antipsychotics and were more than 50% likely to be cost-effective.

Conclusions The primary and sensitivity analyses indicated that conventional antipsychotics may be cost-saving and associated with a gain in QALYs compared with atypical antipsychotics.

Declaration of interest Funding from the Secretary of State for Health (UK). The authors have received consultancy fees, honoraria or departmental support from the pharmaceutical industry, including Novartis, BMS Otsuka, Janssen Cilag, Servier, BMS, Astra-Zeneca, Sanofi-Synthelabo, Eli Lilly and Johnson & Johnson. F.G. is related to an employee of Eli Lilly & Co.

Published economic evaluations suggest that atypical (second-generation) antipsychotics are cost-effective compared with the conventional (first-generation) antipsychotics, chlorpromazine and haloperidol (Almond & O’Donnell, 2000; Gregor et al, 2000; Lewis et al, 2001; Lynch et al, 2001; Martin et al, 2001; Oh et al, 2001; Gianfrancesco et al, 2002; Hosak & Bahbouh, 2002; Palmer et al, 2002; Tilden et al, 2002). However, the available economic and clinical evidence is limited in scale and methodology and the narrow range of antipsychotic drugs considered. Many economic evaluations are modelling studies (to synthesise data from several sources) where it is not possible to verify the source or quality of the data used. Thus, it is not clear that the clinical and economic evidence is sufficient for clinical decision-makers to make treatment choices between the first-generation and second-generation drugs currently available (Davies & Lewis, 2000; Knapp et al, 2002; Bagnall et al, 2003). This is particularly pertinent for people who require a change of antipsychotic but are not eligible for treatment with clozapine.

The aim of our economic evaluation was to inform policy and treatment decisions about the relative costs and utility (or value) of switching treatment between first-generation and second-generation antipsychotics in people with schizophrenia. Specific research questions were, first, are there differences in the direct costs, health state and utility of treatment between first- and second-generation antipsychotics? Second, are first-generation antipsychotics likely to be more cost-effective than second-generation antipsychotics in a population responding poorly to – or intolerant of – current treatment?

METHO

Economic data were collected prospectively for all patients randomised to treatment in an integrated clinical and economic multicentre, open (that is, both clinician and patient knew which drug was being prescribed), rater-blind, randomised controlled trial of alternative classes of antipsychotic drugs (first-generation v. second-generation) in routine National Health Service (NHS) practice in the UK (Jones et al, 2006). The patient population comprised people for whom a change in antipsychotic drug treatment was being considered because of intolerance or insufficient clinical improvement, and for whom a choice between a first-generation antipsychotic and a second-generation antipsychotic other than clozapine was relevant. Inclusion criteria were a DSM-IV diagnosis of schizophrenia, schizoaffective disorder or delusional disorder (American Psychiatric Association, 1994), age 18–65 years and an interval of at least 1 month since the first onset of positive psychotic symptoms. Exclusion criteria were substance misuse or a medical disorder considered clinically to be the major cause of positive psychotic symptoms, or a history of neuroleptic malignant syndrome.

Key characteristics of the trial included: (a) concealed randomisation to the two treatment arms; (b) masked independent assessments of outcome for 1 year following randomisation; (c) intention-to-treat analysis; (d) trial entry defined by the treating clinician deciding to change drug management; (e) broad inclusion criteria to reflect normal clinical practice; (f) choice of drug within a class of treatment was made in advance by the treating clinician; (g) non-commercial funding.

The primary outcome was the score on the Quality of Life Scale (QLS; Heinrichs et al, 1984). The following drugs were available to participants randomised to the first-generation drug treatment arm: chlorpromazine, flupentixol, haloperidol, loxapine, sulpiride, trifluoperazine and zuclopenthixol, plus depot antipsychotics (fluphenazine, zuclopenthixol, flupentixol and haloperidol decanoate). For patients randomised to the second-generation drug treatment arm, the available medications were risperidone, olanzapine, amisulpride and quetiapine. Clinicians were asked to choose the individual drug for their patient
before randomisation. A total of 275 patients were referred (70% of whom were taking first-generation drugs at baseline), 82% (n=227) were randomised and 81% (n=185) of randomised patients completed follow-up. There was 75% power to test the main clinical hypothesis. The overall conclusion was that there was a trend for patients in the first-generation drug treatment arm to do better than those in the second-generation arm, in contrast to the trial hypothesis. There was a difference of 1.7 points on the QLS in favour of first-generation drugs (standard error of difference 1.4; 95% CI –4.5 to 1.1); however, this advantage failed to reach statistical significance (P=0.24).

The economic evaluation used the framework of cost-effectiveness acceptability analysis (Briggs & O’Brien, 2001; Fenwick et al, 2001; Pedram-Sendi & Briggs, 2001; O’Brien & Briggs, 2002) and the perspectives of the NHS, social support services and patients for the primary analysis. These represent the main stakeholders to approximate a broad societal viewpoint or perspective. The analysis included only the direct costs of care, in line with international guidelines and UK policy (Gold et al, 1996; National Institute for Clinical Excellence, 2004). The evaluation was designed to inform policy and treatment decisions in secondary and primary care for a 1-year period, the length of scheduled follow-up from randomisation in the trial. Discounting future costs and outcomes to adjust for time preferences was not necessary for the 1-year time frame.

**Quality-adjusted life-years**

The health measure for the economic evaluation was the quality-adjusted life-year (QALY), calculated from health states reported by all patients enrolled in the trial, using the EuroQoL EQ-5D (Kind, 1996) at baseline and at the 12-week, 26-week and 52-week follow-up assessments. The EQ-5D is a validated generic health status measure covering five domains (mobility, self-care, usual activity, pain/distress, anxiety/depression) and is used in national health surveys in the UK and in clinical trials in mental health. The health status profiles were converted to utility values using published utility tariffs for the EQ-5D (Dolan et al, 1995). The utility values are a measure of preferences for different health states and the relative value of different health states on a scale anchored by death and full health. The utility values were used to estimate QALYs, based on the observed number of days patients were alive in the 12-month follow-up period of the trial.

**Direct costs**

The direct cost of events was estimated from service use observed in the clinical trial multiplied by published national unit cost data (Chartered Institute of Public Finance Accountants, 2002; Netten & Curtis, 2002; Department of Health, 2003). All unit costs were standardised to 2001–2002 prices using a health service price index where necessary (Netten & Curtis, 2002). Service use data were collected at each scheduled follow-up assessment for all patients enrolled in the trial.

Resource use data were collected for hospital in-patient and out-patient services, primary and community care and prescribed medications. First, data on the use of psychiatric hospital care and medication were obtained for all patients, by case-note review in the main psychiatric hospital used by each patient (typically the hospital at which the patient was treated when referred to the trial). Second, patients completed an economic questionnaire at each assessment to identify whether they had used any other hospital, primary or community care services since the previous assessment. Community care included day care facilities and contact with multidisciplinary mental healthcare professionals and teams, social workers and social support workers. If additional services were used, patients were asked to specify the name and location of the services. Third, additional data on the number of times each service was used (as identified by the patient in the economic questionnaire) were obtained from detailed review of the relevant clinical records for each person. These three methods of data collection minimised the extent of missing data for key cost drivers (psychiatric in-patient and out-patient hospital care).

National average unit cost data were used to control for differences in costs between care settings. The national reference cost data published by the UK Department of Health (2003) were used to estimate the cost of psychiatric in-patient and out-patient care, by type of ward or outpatient visit. Sensitivity analysis was used to test the impact of using national unit cost data from other sources for psychiatric hospital costs (Chartered Institute of Public Finance Accountants, 2002; Netten & Curtis, 2002). The hospital trust financial returns data published by the Chartered Institute of Public Finance Accountants (CIPFA) were used to estimate the cost of non-psychiatric hospital care by type of ward or admitting specialty. The reference cost data-set did not have detailed unit costs for non-psychiatric hospitals, so the more detailed CIPFA database was used for these costs.

Information was collected for each patient about dosage, duration and route of administration of medication. A daily cost for oral medication and cost per injection or dose for depot and pro re nata medicines was estimated by multiplying the quantity of medication by unit costs derived from the British National Formulary (British Medical Association & Royal Pharmaceutical Society of Great Britain, 2002). The daily cost was multiplied by the reported duration of courses of treatment completed within the study period, and by the length of the study period for continuing courses of treatment. The cost of medicines did not include the costs of dispensing or administration of the drugs (e.g. by injection), as it was assumed that these were included in the unit costs used for hospital in-patient and out-patient care and primary and community care.

**Incremental cost-effectiveness analysis**

Descriptive analysis was used to compare utility values, QALYs and costs. The primary measure of interest for the economic analysis was the incremental cost-effectiveness ratio (ICER). The QALY and cost data were used as inputs to estimate the ICER and cost-effectiveness acceptability curves, meaning that statistical analysis of these was not necessary. Incremental cost-effectiveness ratios were calculated as the difference in costs divided by the difference in QALYs for the two types of medication (Gold et al, 1996). The incremental costs and QALYs were estimated by analysis of covariance (ANCOVA) using a general linear model and covariates of baseline QLS score, utility, psychiatric hospital in-patient and outpatient costs prior to enrolment in the trial and trial centre (location). Treatment allocation was entered as a fixed factor.

Statistical measures of variance of the ICER were not calculated, since standard methods of analysis do not allow these to be estimated in any meaningful way.
The net benefit statistic gives an estimate of willingness to pay to gain a QALY.

where the net benefit of an intervention is the proportion of bootstrapped simulations where the net benefit of an intervention is greater than zero for each WTPT (Fenwick et al., 2001, Pedram-Sendi & Briggs, 2001; Hoch et al., 2002; O’Brien & Briggs, 2002). The WTPT values used ranged from decision-makers being willing to pay £1 to gain 1 QALY to decision-makers being willing to pay £50,000 to gain 1 QALY. This includes the range of implied values that are acceptable to policy-makers in the UK (Rawlins & Culyer, 2004).

The cost-effectiveness acceptability curve summarises the information at each value of willingness to pay to gain a QALY. The net benefit statistic gives an estimate of the monetary value of a QALY or other measure of effectiveness.

Data manipulation and missing data

The economic data were manipulated and analysed using the Statistical Package for the Social Sciences version 11.5 for Windows to calculate costs and QALYs for the 1-year follow-up period and to estimate missing utility data. Missing utility data for patients who completed the scheduled follow-up but had missing observations were imputed by linear interpolation (value of previous period plus value of next period divided by 2), if observations either side of the missing item were available. Patients with one or more missing observations at the end of follow-up were treated as censored cases due to withdrawal or loss to follow-up. The survival function and probability of survival at each assessment point was estimated, using patient status (alive, dead or withdrawn) and treatment allocation. Multiple imputation (propensity score) was used to impute values for the missing costs, by category of resource use using SOLAS for Missing Data Analysis version 3.0 (Statistical Solutions Ltd, Cork, Ireland) (Rubin & Schenker, 1991; Lavori et al., 1995). This meant that missing cost data were treated as missing at random, rather than informative censoring of data. This was based on the assumption that use of services and subsequent costs were determined by a range of factors in addition to treatment allocation or previous service use. Analysis of covariance and the bootstrap analysis were conducted in Stata version 9 for Windows.

Sensitivity analyses

Some assumptions were required to deal with missing data. The impact of these assumptions on the results was tested using alternative approaches to imputation of missing data. The impact of alternative sources of unit cost data was also tested in the sensitivity analysis.

RESULTS

The clinical and demographic characteristics of patients in the two randomised groups were similar at baseline (Jones et al., 2006), and differences in health status, utility and costs for the previous 3 months were not statistically significant (Table 1).

Quality-adjusted life-years

The unadjusted health status and utility scores at baseline and 12-month follow-up assessments indicate that the health-related quality of life of participants improved over the 12 months from baseline (Tables 1 and 2). There was an observed difference in QALYs, including imputed values for missing data (Table 2), which at least partially reflects differences in utility at baseline. Differences between the groups in adjusted utility values are highest at weeks 12 and 26, and diminish by week 52 (Fig. 1).

Costs

Participants in both treatment groups mostly used psychiatric hospital services (Table 3). Data on the use of psychiatric hospital care at 52 weeks were available for a high proportion of patients: 88% in the first-generation antipsychotic (FGA) group and 91% in the second-generation antipsychotic (SGA) group. Data were less complete for other cost categories, the lowest rate of follow-up being the use of primary and community care services at 77%. Overall, 85% of patients reported that they had used primary and community care services and their records were reviewed to identify intensity of resource use. Total cost data were available for 65% of participants. The use of psychiatric hospital care constituted 88% of the total costs of care (91% FGA, 84% SGA).

There was a trend for the mean costs (including imputed costs for missing observations and censored cases) to be lower for people allocated to the FGA group than the SGA group (Table 4). The costs of care at each follow-up period and over 1 year are characterised by large standard deviations, reflecting large differences between patients in the use of services. The costs of antipsychotic medication were a small percentage of overall costs (2% FGA and 4% SGA).

Cost-effectiveness analysis

The primary and sensitivity analyses indicated that switching therapy to a first-generation antipsychotic may result in lower costs and higher QALYs (Table 5). All of the primary and sensitivity analyses indicate a large standard error associated with the differences in costs and QALYs. This indicates a high level of variation in
these variables between patients, and therefore a high level of uncertainty associated with the estimated differences in costs and QALYs.

Figure 2 presents the probability that first-generation antipsychotics are cost-effective in the form of a cost-effectiveness acceptability curve. If decision makers were willing to pay up to £35 000 to gain 1 QALY, then the probability that these drugs are cost-effective is 0.75, with an associated net benefit of £1752. That is, 75% of the pairs of bootstrap replicates indicated that these drugs were associated with a net benefit value greater than zero (the net cost of the drug minus the net QALY multiplied by £35 000). The ceiling cost per QALY ratio of £35 000 is at the top of the range of implied values that are acceptable to policy-makers in the UK (Rawlins & Culyer, 2004).

Overall, the probability that first-generation antipsychotics are cost-effective is between 0.54 (if decision-makers were willing to pay only £1 to gain 1 QALY) and 0.81 (if decision-makers were willing to pay up to £50 000 to gain 1 QALY).

**DISCUSSION**

The primary and sensitivity analyses suggest that first-generation antipsychotic drugs may be associated with small cost savings and a small gain in QALYs when compared with second-generation drugs. The cost-effectiveness acceptability analysis supported this conclusion. These results were estimated from resource use and health status data collected as an integral part of a randomised controlled clinical trial. The use of a randomised controlled trial strengthens the reliability and internal validity of the data collected; however, aspects of the trial design may affect the validity or robustness of the data.

**Potential limitations**

Sequential statistical tests of differences in costs and QALYs were not conducted. This was because the primary measure of outcome for the economic analysis was the incremental cost-effectiveness ratio. As a ratio, this is not amenable to statistical analysis of differences between groups (Briggs et al., 2002). An alternative approach to assessing the level of variance and uncertainty associated with the data was to estimate net benefit statistics and cost-effectiveness acceptability curves. The latter estimate the likelihood or probability that first-generation drugs are more or less cost-effective than second-generation ones. Other advantages of this approach are, first, that the QALY and cost data were inputs to estimate the ICER. If cost and outcome interact (i.e. poorer health status is associated with increased resource use and cost), then it is more appropriate to relate net costs to patient outcomes than sequentially test for statistical differences in costs and QALYs. Cost-effectiveness acceptability curves incorporate this interaction between costs and QALYs and provide a method to assess the uncertainty associated with the data. Second,
the trial was not powered to detect differences in costs, QALYs or net benefit. Insufficient power increases the chance of type II errors (failing to reject the null hypothesis of no difference between groups when a difference does exist). Post hoc sample size calculations indicate that the power to detect statistically significant differences in net benefit was low. If decision-makers consider important differences in costs and QALYS to be £1500 and 0.10 respectively, and are prepared to pay £35000 to gain 1 QALY, then there was 25% power to detect statistically significant differences in net benefit. If decision-makers consider smaller differences in costs and QALYS to be important, or are prepared to pay less to gain 1 QALY, then the power to detect statistically significant differences in net benefit was lower.

Both the participants and referring clinicians knew of the treatment allocation and drug prescribed, so subjective patient responses to the EQ–5D and service use measures might have been influenced by knowledge of treatment allocation. In addition, knowledge of the treatment allocation

---

### Table 2  Health status, length of follow-up, utility and QALY scores at 1 year, unadjusted for baseline covariates

<table>
<thead>
<tr>
<th>Health status, n (%)</th>
<th>FGA (n=118)</th>
<th>SGA (n=109)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility (walking)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No problems</td>
<td>80 (68)</td>
<td>64 (59)</td>
</tr>
<tr>
<td>Missing</td>
<td>20 (17)</td>
<td>25 (23)</td>
</tr>
<tr>
<td>Self-care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No problems</td>
<td>89 (75)</td>
<td>67 (61)</td>
</tr>
<tr>
<td>Missing</td>
<td>20 (17)</td>
<td>25 (23)</td>
</tr>
<tr>
<td>Usual activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No problems</td>
<td>56 (47)</td>
<td>43 (39)</td>
</tr>
<tr>
<td>Missing</td>
<td>20 (17)</td>
<td>25 (23)</td>
</tr>
<tr>
<td>Pain/discomfort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No pain or discomfort</td>
<td>77 (65)</td>
<td>61 (56)</td>
</tr>
<tr>
<td>Missing</td>
<td>20 (17)</td>
<td>25 (23)</td>
</tr>
<tr>
<td>Anxiety/depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not anxious or depressed</td>
<td>44 (37)</td>
<td>38 (35)</td>
</tr>
<tr>
<td>Missing</td>
<td>20 (17)</td>
<td>25 (23)</td>
</tr>
<tr>
<td>Follow-up, days: mean (s.d.)</td>
<td>354 (85)</td>
<td>344 (93)</td>
</tr>
<tr>
<td>Utility value: mean (s.d.)</td>
<td>0.78 (0.22)</td>
<td>0.75 (0.23)</td>
</tr>
<tr>
<td>QALY: mean (s.d.)</td>
<td>0.74 (0.22)</td>
<td>0.67 (0.25)</td>
</tr>
</tbody>
</table>

FGA, first-generation antipsychotics; QALY, quality-adjusted life-year; SGA, second-generation antipsychotics.

---

### Table 3  Use of services: comparison of the two treatment groups

<table>
<thead>
<tr>
<th></th>
<th>Weeks 0–12</th>
<th>Weeks 13–26</th>
<th>Weeks 27–52</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FGA</td>
<td>SGA</td>
<td>FGA</td>
</tr>
<tr>
<td><strong>Psychiatric hospital</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>115</td>
<td>106</td>
<td>111</td>
</tr>
<tr>
<td>In-patient days: mean (s.d.)</td>
<td>29 (37)</td>
<td>28 (36)</td>
<td>20 (36)</td>
</tr>
<tr>
<td>Day patient and out-patient visits: mean (s.d.)</td>
<td>0 (0)</td>
<td>2 (13)</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Non-psychiatric hospital</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>112</td>
<td>97</td>
<td>108</td>
</tr>
<tr>
<td>In-patient days: mean (s.d.)</td>
<td>1 (10)</td>
<td>3 (29)</td>
<td>0 (2)</td>
</tr>
<tr>
<td>Out-patient visits: mean (s.d.)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Community and primary care</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>90</td>
<td>83</td>
<td>86</td>
</tr>
<tr>
<td>Psychiatry and psychology visits: mean (s.d.)</td>
<td>0 (1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>GP, district and practice nurse visits: mean (s.d.)</td>
<td>1 (2)</td>
<td>2 (2)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Other primary and community care staff visits: mean (s.d.)</td>
<td>1 (6)</td>
<td>1 (3)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Community centre/day centre visits: mean (s.d.)</td>
<td>3 (8)</td>
<td>5 (14)</td>
<td>4 (12)</td>
</tr>
<tr>
<td><strong>Antipsychotic medication</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>111</td>
<td>104</td>
<td>105</td>
</tr>
<tr>
<td>Days per antipsychotic: mean (s.d.)</td>
<td>60 (24)</td>
<td>53 (25)</td>
<td>73 (32)</td>
</tr>
<tr>
<td>Number of antipsychotics: mean (s.d.)</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td>2 (1)</td>
</tr>
<tr>
<td><strong>Non-antipsychotic medication</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>91</td>
<td>75</td>
<td>88</td>
</tr>
<tr>
<td>Days per medicine: mean (s.d.)</td>
<td>74 (28)</td>
<td>69 (26)</td>
<td>84 (29)</td>
</tr>
<tr>
<td>Number of medicines: mean (s.d.)</td>
<td>2 (1)</td>
<td>2 (2)</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

FGA, first-generation antipsychotics; GP, general practitioner; SGA, second-generation antipsychotics.
and drug prescribed might have influenced the assessment and interpretation of effectiveness and side-effects by the referring clinician and participant (Lloyd et al., 2005). This may in turn have influenced treatment decisions and subsequent service use and outcomes of the patients. This is particularly important if treating clinicians had prior expectations of the likely effectiveness and side-effects of particular antipsychotics or classes of antipsychotic. However, it might be expected that this would operate in favour of the newer, second-generation antipsychotics rather than the older conventional drugs. In addition, knowledge of the drug prescribed is an important factor that may increase the general applicability of the results.

As part of the operational protocol, clinicians were asked to try to keep participating patients on the randomised medication for at least the first 12 weeks (if compatible with good practice). If the medication needed to be changed, the clinician was asked to prescribe another drug within the same class, if possible. If changes to medication occur more frequently in routine practice, the costs estimated in this trial may be underestimated. Changes of medication could incur additional costs for hospital in-patient and out-patient care. At 52 weeks there was a trend for more participants to remain in the SGA arm (65%) than the FGA arm (51%). This is reflected by the increasing cost of antipsychotic medication in the FGA arm over the course of the 1-year follow-up.

The EQ-5D, an instrument shown to have acceptable validity in people with schizophrenia in European countries (Prieto et al., 2003; Bobes, 2003; Bobes, 2004), may not be sensitive to small but important changes in the symptoms and health-related quality of life of people with schizophrenia. The QALY and the EQ-5D have not been widely used in mental healthcare studies in the past, but the need to demonstrate value for money of interventions and comply with international standards and local guidelines for the design of economic evaluations supports their use (Gold et al., 1996; National Institute of Clinical Excellence, 2004). The differences in QALYs in this study were small, but similar to those

### Table 4 Costs of services

<table>
<thead>
<tr>
<th>Costs, £: mean (s.d.)</th>
<th>Weeks 0–12</th>
<th>Weeks 13–26</th>
<th>Weeks 27–52</th>
<th>Annual cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGA</td>
<td>SGA</td>
<td>FGA</td>
<td>SGA</td>
<td>FGA</td>
</tr>
<tr>
<td>(n=118)</td>
<td>(n=109)</td>
<td>(n=118)</td>
<td>(n=109)</td>
<td>(n=118)</td>
</tr>
<tr>
<td>Psychiatric hospital</td>
<td>5728 (7885)</td>
<td>5587 (7296)</td>
<td>4372 (8161)</td>
<td>4256 (7471)</td>
</tr>
<tr>
<td>Non-psychiatric hospital</td>
<td>177 (1633)</td>
<td>243 (1670)</td>
<td>118 (575)</td>
<td>329 (2459)</td>
</tr>
<tr>
<td>Antipsychotic medicines</td>
<td>73 (114)</td>
<td>179 (174)</td>
<td>70 (83)</td>
<td>200 (211)</td>
</tr>
<tr>
<td>Other medicines</td>
<td>41 (69)</td>
<td>36 (50)</td>
<td>44 (80)</td>
<td>43 (61)</td>
</tr>
<tr>
<td>Community and primary care</td>
<td>181 (236)</td>
<td>248 (386)</td>
<td>230 (347)</td>
<td>291 (632)</td>
</tr>
<tr>
<td>Total cost</td>
<td>6200 (7947)</td>
<td>6292 (7350)</td>
<td>4835 (8150)</td>
<td>5119 (6787)</td>
</tr>
</tbody>
</table>

FGA, first-generation antipsychotics; SGA, second-generation antipsychotics.

### Table 5 Incremental costs and quality-adjusted life years (QALYs) and cost per QALY, adjusted for covariance

<table>
<thead>
<tr>
<th></th>
<th>Net cost, £ (s.e.)</th>
<th>Net QALY (s.e.)</th>
<th>Net monetary benefit if WTP=£35 000, £</th>
<th>FGA cost-effective if WTP=£35 000, % simulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary analysis</td>
<td>–116 (2464)</td>
<td>0.04 (0.03)</td>
<td>1752</td>
<td>75</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis unadjusted for covariance</td>
<td>–1260 (3599)</td>
<td>0.07 (0.03)</td>
<td>3720</td>
<td>84</td>
</tr>
<tr>
<td>Complete case analysis</td>
<td>–2390 (2737)</td>
<td>0.02 (0.02)</td>
<td>3308</td>
<td>87</td>
</tr>
<tr>
<td>QALYs including imputed values for missing observations only</td>
<td>–116 (2391)</td>
<td>0.04 (0.02)</td>
<td>1650</td>
<td>74</td>
</tr>
<tr>
<td>Net costs, PSSRU unit cost data</td>
<td>–1157 (2248)</td>
<td>0.04 (0.03)</td>
<td>2750</td>
<td>88</td>
</tr>
<tr>
<td>Net costs, CIPFA unit cost data</td>
<td>–1418 (2230)</td>
<td>0.04 (0.03)</td>
<td>3008</td>
<td>90</td>
</tr>
</tbody>
</table>

CIPFA, Chartered Institute of Public Finance Accountants; FGA, first-generation antipsychotics; PSSRU, Personal Social Services Research Unit; WTP, willingness to pay to gain 1 additional QALY.

found in UK-based modelling studies using QALYs (Davies & Lewis, 2000; Bagnall et al., 2003). Changes in health status and in utility were detected over the 52-week follow-up period. Additionally, there was a statistically significant correlation between the utility values and other measures used in the trial, including the primary clinical outcome measure, the QLS.

The costs of contacts with the criminal justice system, use of residential accommodation and informal care were excluded, as were the indirect costs of withdrawal from paid employment. Thus, the total costs were underestimated, which might have biased the results if there were important differences in utilisation due to the choice of antipsychotic rather than the influence of organisational and social factors. However, a descriptive analysis of these variables suggests that the level of use was low and that there were few differences in the use of these services over the 12-month period of the trial.

The economic analyses included multiple imputation techniques to generate values for all missing observations and censored cases, to reduce the impact of bias due to attrition. The ANCOVA supported this approach to missing data. The use of imputation reduced the cost difference between the two types of antipsychotics when compared with the complete case analysis. In contrast, the imputation of missing utility data increased the difference between the two drug classes. Using the imputed and complete case data-sets, the results for patients randomised to switch to first-generation antipsychotics were associated with a trend towards higher QALYs and lower costs than those of patients randomised to receive second-generation antipsychotics. The analysis using imputed data reduces the potential for bias associated with missing data, which is particularly important in studies with multiple follow-up points.

Overall total cost data were available for 65% of participants. However, the use of psychiatric hospital care comprised 88% of the total costs of care observed. Data on this key cost driver were available for a high proportion of patients randomised to treatment (88% FGA and 91% SGA). This suggests that the impact of missing data on total cost per person due to attrition is likely to be relatively low.

The demographic and clinical characteristics of participants at baseline were well balanced between the treatment allocation groups for most characteristics. However, there were observed differences between the groups in health status and utility as measured by the EQ-5D, the costs of hospital care for the 3 months prior to randomisation, and the trial centre in which the patient was treated. These factors were included as covariates in the cost-effectiveness analyses. The analysis of covariance reduced the benefit of first-generation antipsychotics compared with an unadjusted analysis, but still indicated that these drugs are likely to be more cost-effective than second-generation antipsychotics.

Although the randomisation procedure appeared to result in well-balanced groups, the participants referred to the trial might have been a selected and unrepresentative sample of patients. Parallel audit in the clinical services in two of the centres suggested that only 20–37% of possibly eligible patients (those with a diagnosis of schizophrenia whose drug treatment was being changed) were randomised into the trial. The remaining patients were either not referred to the trial or refused to participate. There was insufficient information to determine whether the patients who participated in the trial were representative of eligible patients requiring a change in medication. Therefore, patients referred into the trial might not be representative of the population of patients requiring a change in medication owing to poor response or intolerance, reducing generalisability to the population of interest. However, there was an apparent lack of equipoise for both clinicians and patients, with a belief that second-generation antipsychotics were superior to the first-generation drugs. A survey of the attitudes of clinicians at each of the trial centres, conducted as part of the trial, supports this lack of clinical equipoise. The survey found that 90% of respondents believed that second-generation antipsychotics were associated with less severe side-effects than the conventional drugs and 38% believed that the former were superior in terms of clinical efficacy (Lloyd et al., 2005). If the main reason for the low participation rate was a belief in the superiority of second-generation antipsychotics, then the trial sample may be representative of the eligible population.

The trial was conducted in five centres in England, covering 14 NHS trusts. The trusts and trial centres represented a range of geographical areas, with populations that varied in socio-demographic and economic characteristics; the trial settings thus are likely to include the range of treatment settings and patient populations encountered in routine practice.

The trial followed patients for 12 months from baseline. Although this is a relatively long period of follow-up compared with earlier trials of antipsychotic medication, it may not be long enough to observe changes in costs and outcomes over the course of a chronic illness. The data indicate that over the course of the 12-month follow-up period up to half of patients changed medication. In addition, both health status and costs changed over this
time. Therefore, the results of this analysis may not reflect what will happen over the longer term.

Comparison with previous studies
The results of this analysis accord with the overall conclusions of two UK-based economic modelling studies (Davies & Lewis, 2000, Bagnall et al, 2003). The majority of economic studies comparing first- and second-generation antipsychotics suggest that the latter drugs may be cost-effective (Almond & O'Donnell, 2000; Gregor et al, 2000; Lewis et al, 2001; Lynch et al, 2001; Oh et al, 2001; Gianfrancesco et al, 2002; Hosak & Bahbouh, 2002; Palmer et al, 2002; Tilden et al, 2002). However, the robustness of these studies is uncertain since they are limited in the range of antipsychotic drugs considered, scale and methodology (Davies & Lewis, 2000; Knapp et al, 2002; Bagnall et al, 2003). For instance, the usual first-generation antipsychotic comparator in previous studies was haloperidol, known to cause substantial rates of side-effects (Geddes et al, 2000). In addition, most of these studies rely on simulations of data from short-term efficacy trials with no primary economic focus. In contrast, in this study haloperidol was selected by clinicians in only 8% of cases randomised to the FGA arm.

Implications of the study
Overall, this study confirms that there is no evidence to suggest second-generation antipsychotics are more cost-effective than first-generation ones. This is supported by recent meta-analyses of the clinical evidence (Lieberman, 2006), an observational study (Kilian et al, 2004) and two pragmatic, long-term, randomised studies (Rosenheck et al, 2003; Lieberman et al, 2005), which failed to find evidence for the superiority of second-generation antipsychotics in terms of effectiveness or quality of life.

Practice and prescribing guidelines categorise the available antipsychotics for schizophrenia into first- and second-generation. This study is the first, non-commercially funded, integrated economic and clinical trial to reflect routine practice and guidelines and compare the two classes of antipsychotics in the context of the NHS. The primary and sensitivity analyses of the economic data indicate that first-generation antipsychotics may be cost-saving and associated with a gain in QALYs compared with the second-generation agents. There was no evidence that the second-generation drugs were more cost-effective than the first-generation ones. The cost-effectiveness acceptability analysis supported this conclusion. In other words, for people who required a change in treatment, switching to a first-generation antipsychotic may be as – or more – cost-effective than switching to a second-generation antipsychotic. However, as described above, there were limitations to the study that increase the level of uncertainty in the economic results.

None the less, the data add to a growing body of evidence that questions the perception that second-generation antipsychotics are superior to the earlier drugs in terms of clinical effectiveness, quality of life and cost-effectiveness. The data from this large, pragmatic, randomised controlled trial suggest that careful prescribing of first-generation antipsychotics in routine practice may be cost-effective. Further observational and pragmatic trials are required to identify cost-effective anti-psychotic use, the determinants of costs and outcomes and the roles of first- and second-generation antipsychotic drugs in long-term management.

ACKNOWLEDGEMENTS

Financial support was provided by a project grant from the Secretary of State for Health under the UK NHS Technology Assessment Programme (project number 96/1906). The views and opinions expressed by the authors do not necessarily reflect those of the Department of Health.

In addition to the authors, members of the CUTHALL team included Graham Dunn, Robin Murray, Rob Kerwin, David Taylor and Martin Knapp who contributed to the design and analysis of the clinical and economic trial. Maria Clark, Tanya Hawthorn, Fiona Hynes, Fiona Bar, Leilani, Jenny Massie, Ahmed Mahmoud, Paul Monks, Alex Barrow, Tracy Fay, Maurice Gervin and Susie Morrow were essential in the management and execution of the study. Patient assessments were carried out by Candice Blackwell, Nerjis Gooding, Ronina Howitt, Natasha Newbery, Eleanor Page, Joanne Shepherd and Emma Sowden. Secretarial support from Patricia Smith and Helen Woodiwiss. Data entry and IT support from Simon Foster, Xinning Jin, Zhenhua Zhu, John Cooley, Paul Schofield and Lisa Riley.

REFERENCES


Metabolic disease and cardiovascular risk in people treated with antipsychotics in the community

PAUL MACKIN, DAVID BISHOP, HELEN WATKINSON, PETER GALLAGHER and I. NICOL FERRIER

Background Prevalence of physical comorbidity in severe mental illness is a significant public health concern, but comparative data in people with diagnoses other than schizophrenia are sparse.

Aims To investigate the prevalence of metabolic disease and cardiovascular risk in people with severe mental illness treated with antipsychotics in the community.

Methods Case–control study of 90 people treated with antipsychotics in the community and 92 age- and gender-matched controls. The prevalence of metabolic syndrome and 10-year cardiovascular risk were calculated.

Results People on antipsychotics had a significantly worse metabolic profile than controls (F = 6.583, d.f. = 15,161, P < 0.0001). Moreover, metabolic syndrome was more prevalent (OR = 3.68, 95% CI 1.71–7.93, P = 0.001), as was cardiovascular risk across a number of outcomes. These results are consistent across diagnostic groups.

Conclusions People with severe mental illness treated with antipsychotics have excess metabolic dysfunction and heightened risk for cardiovascular disease.

Declararion of Interest PM, I.N.F. and PG. have received honoraria for educational meetings from pharmaceutical companies. Funding detailed in Acknowledgements.

Severe mental illness is associated with a significant excess of physical comorbidity and mortality (Brown 1997; Phelan et al., 2001; Osborn et al., 2007), and as such represents a major public health concern. Although suicide is prevalent in this population, ischaemic heart disease, not suicide, may be the major contributor to excess mortality (Lawrence et al., 2003). Recently published UK guidelines on the management of schizophrenia (National Institute for Clinical Excellence, 2002) and bipolar disorder (National Institute for Clinical Excellence, 2006) recognize the impact of physical comorbidity in these disorders, as well as the paucity of high-quality research in this field.

A number of recent studies have quantified the risk of coronary heart disease, based on Framingham risk estimates, in people with severe mental illness (Goff et al., 2005; Correll et al., 2006; Osborn et al., 2006), but these focused only on those with a diagnosis of schizophrenia or non-affective psychoses (Goff et al., 2005; Osborn et al., 2006) and hospital in-patients (Correll et al., 2006). In this study we determined the prevalence of metabolic dysfunction and estimates of cardiovascular risk in a community sample from secondary care of people with severe mental illness from across the diagnostic spectrum, who were taking antipsychotics, and compared the results with those from age- and gender-matched controls.

METHOD

Participants Patients from all secondary care community mental health services from across the former Newcastle, North Tyneside and Northumberland Mental Health NHS Trust, and the Regional Affective Disorders Tertiary Service were invited to participate in a baseline study of metabolic dysfunction between January 2002 and March 2004. Participants were recruited irrespective of psychiatric diagnosis. Inclusion criteria were a psychiatric diagnosis and the prescription of and adherence to (determined by self-report) antipsychotic medication for a minimum of 6 months. People with a known diagnosis of type 1 or type 2 diabetes mellitus, anorexia nervosa, bulimia nervosa, neoplastic disease or alcohol dependence were excluded. We invited 198 people to participate and 106 (54%) gave their informed consent. Baseline characteristics of this cohort have been described previously (Mackin et al., 2003).

All participants with a baseline assessment of metabolic function were invited to participate in a follow-up study between June and December 2005. An age- and gender-matched control group was recruited between January and June 2006 for comparison of metabolic and cardiovascular risk parameters. In an attempt to control for demographic and socio-economic variables, family members and carers were invited to participate as controls, and advertisements for volunteers were placed in local facilities within the geographical environs in which the community mental health teams were based. People with a history of psychiatric disorder and those who had ever taken a prescribed drug for a psychiatric disorder were excluded. All participants gave written informed consent and the study was approved by the Newcastle local research ethics committee.

Procedures Participants were given written instructions to fast overnight on the day before assessment, and were asked to confirm their fasting status on the morning of study. All assessments were performed in the Department of Psychiatry, University of Newcastle upon Tyne between 08.30 and 10.00 h on the study day. Demographic details of age, gender and ethnic group were obtained. Current and previous tobacco, alcohol and illicit substance use were recorded, as well as any history of cardiovascular disease and diabetes mellitus in first-degree relatives. Information regarding psychiatric diagnosis, duration of illness, number of admissions to psychiatric in-patients facilities, medication (including non-psychotropic drugs) and dosage was recorded and confirmed, where necessary, by reference to case notes and general practitioner records.
Height, weight, and waist and hip circumference were recorded using standardised procedures. Body mass index (BMI) and waist-to-hip ratio were calculated. Conventional BMI categories were used (underweight <18.5; normal 18.5–24.9; overweight 25.0–29.9; obese: >30.0). Blood pressure was recorded using a sphygmomanometer on three occasions during the assessment, and the value expressed as the mean of the three recordings. A 12-lead electrocardiogram (ECG) was recorded at 50 mm/s using a MAC 1200ST portable machine (GE Medical Systems, Slough, Berkshire, UK). For the purposes of cardiovascular risk estimation, ECGs were analysed for Framingham voltage criteria for left ventricular hypertrophy (Levy et al, 1990).

A single venous blood sample was withdrawn and analysed for glucose, glycosylated haemoglobin (HbA1c), insulin and lipid profile (total cholesterol, high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterol, and triglycerides). Insulin was measured by enzyme-linked immunosorbent assay. The Homeostasis Model Assessment (HOMA; Levy et al, 1998) was used to assess glucose handling, which is expressed as pancreatic beta-cell function, insulin sensitivity and insulin resistance. Values for these parameters were based on fasting glucose and insulin levels and calculated using the HOMA Calculator, version 2.2 (Diabetes Trial Unit, University of Oxford, UK). The model is calibrated to give beta-cell function and insulin sensitivity of 100% in healthy adults with currently available insulin assays. Impaired fasting glucose was defined as fasting blood glucose between 6.1 and 7.0 mmol/l, and diabetes mellitus as fasting blood glucose ≥7.0 mmol/l (National Diabetes Data Group, 1979). The presence of the metabolic syndrome was based on the definition by the International Diabetes Federation (Alberti et al, 2006).

Cardiovascular risk estimates were based on established risk factors using the Joint British Societies’ (JBS) definition of cardiovascular disease, and the Framingham definition (Anderson et al, 1991). The University of Edinburgh Cardiovascular Risk Calculator (http://cvrisk.mvm.ed.ac.uk/calculator.htm) was used to compute percentage risk estimates for a number of outcomes over a 10-year period. Risk estimates using the Framingham equation have important differences from the JBS definition which include the ability to calculate specific risks (for cardiovascular disease, coronary heart disease, myocardial infarction, stroke, death due to cardiovascular disease and death due to coronary heart disease) and the option to vary the time period over which risk is computed. Cardiovascular risk is calculated from the following parameters: age, gender, smoking status, blood pressure, total cholesterol and HDL cholesterol. The Framingham equation also incorporates the presence of left ventricular hypertrophy in the risk estimate.

Statistical analysis

Data were analysed using the Statistical Package for the Social Sciences, version 11 for Windows. Demographic characteristics were examined by t-test or χ² test where appropriate. Owing to the number of metabolic parameters measured and the risk of Type 1 error, we first conducted a multivariate analysis of covariance (MANCOVA) to test for a significant overall difference in continuous metabolic parameters between the group with mental illness and controls. Differences in individual measures were then examined by follow-up t-tests or Mann–Whitney tests.

χ² analysis was used to compare the distribution of discrete variables. Analysis of variance (ANOVA) was used to examine the effect of specific factors such as smoking status or antipsychotic drug (i.e. typical or atypical) on metabolic and cardiovascular risk estimates. All reported P values are two-tailed. Statistical significance is defined as P < 0.05.

RESULTS

Characteristics of participants

Of the original 106 participants in the baseline study, 90 (85%) consented to participate in the current study; 6 (5.7%) did not reply to the invitation; 6 (5.7%) refused to consent; 2 (1.9%) were too unwell to participate; and 1 (1%) denied having participated in the original study. Characteristics of the participants with mental illness and the 92 controls are given in Table 1. The groups were well matched in terms of age and gender. Participants with mental illness were recruited from across the diagnostic spectrum: bipolar disorder (n=32, 35.6%); schizophrenia (n=27, 30.0%); schizoaffective disorder (n=9, 10.0%); other (including delusional,

### Table 1  Characteristics of participants with severe mental illness and controls

<table>
<thead>
<tr>
<th></th>
<th>People with mental illness (n=90)</th>
<th>Controls (n=92)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years: mean (s.d.)</td>
<td>45.7 (11.8)</td>
<td>43.5 (13.6)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>44 (48.9)</td>
<td>43 (46.7)</td>
</tr>
<tr>
<td>Female</td>
<td>46 (51.1)</td>
<td>49 (53.3)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>88 (97.8)</td>
<td>90 (97.8)</td>
</tr>
<tr>
<td>Asian (Indian)</td>
<td>2 (2.2)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Asian (Oriental)</td>
<td>0 (0)</td>
<td>1 (1.1)</td>
</tr>
<tr>
<td>Mixed race</td>
<td>0 (0)</td>
<td>1 (1.1)</td>
</tr>
<tr>
<td>Current smoker, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>36 (40.0)**</td>
<td>13 (14.1)</td>
</tr>
<tr>
<td>No</td>
<td>54 (60.0)</td>
<td>79 (85.9)</td>
</tr>
<tr>
<td>History of substance misuse, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>27 (30.0)**</td>
<td>4 (4.3)</td>
</tr>
<tr>
<td>No</td>
<td>63 (70.0)</td>
<td>88 (95.7)</td>
</tr>
<tr>
<td>Family history of cardiovascular disease, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>61 (67.8)</td>
<td>50 (54.3)</td>
</tr>
<tr>
<td>No</td>
<td>29 (32.2)</td>
<td>42 (45.7)</td>
</tr>
<tr>
<td>Family history of diabetes mellitus, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>31 (34.4)</td>
<td>22 (23.9)</td>
</tr>
<tr>
<td>No</td>
<td>59 (65.6)</td>
<td>70 (76.1)</td>
</tr>
</tbody>
</table>

**P < 0.001.
currently smoked tobacco (40.0 v. 14.1% of controls, \( \chi^2 = 15.87, P < 0.001 \)) and had a history of substance misuse (30.0 v. 4.3% of controls, \( \chi^2 = 21.18, P < 0.001 \)).

### Medication

Of the 90 people with mental illness who participated in this study, 83 (92%) were still receiving antipsychotic medication; 71 (86%) were receiving one antipsychotic drug and 12 (14%) were prescribed combination antipsychotic medication. Of those taking just one antipsychotic, 16 (23%) were taking a typical agent and 55 (77%) an atypical. Details of antipsychotic and other medication are given in Table 2.

### Metabolic parameters

Metabolic parameters for participants with mental illness and controls are given in Table 3. From the MANCOVA model, age was found to be a highly significant covariate (\( F = 5.873, \text{d.f.} = 15,161, P < 0.0001 \)) but those with mental illness had a significantly worse metabolic profile (age-adjusted main effect: \( F = 6.583, \text{d.f.} = 15,161, P < 0.0001 \)). Body mass index, waist circumference, waist-to-hip ratio, total cholesterol, LDL cholesterol, serum triglycerides, fasting blood glucose, HbA1c, and serum insulin were all significantly higher in those with mental illness than controls. Moreover, HDL cholesterol (which is cardioprotective) was significantly lower. Estimation of insulin sensitivity and insulin resistance by HOMA revealed differences between the two groups; that is people with mental illness were more insulin resistant, more had disorders of glucose homeostasis compared with controls (14.4 v. 1.1%, \( P = 0.003 \)), and there was a higher prevalence of the metabolic syndrome (33.3 v. 11.9%, \( P = 0.001 \)). There were no differences in either systolic or diastolic blood pressure between the groups.

### Cardiovascular risk

Ten-year risk estimates based on the JBS definition of cardiovascular disease and the Framingham cardiovascular outcome risk estimates are given in Table 4. The risk calculator allows estimation of risk for people between 35 and 75 years of age (participants with mental illness \( n = 72 \); controls \( n = 65 \)). Figure 1 represents the differences in cardiovascular outcome risks between the two groups.

### Effect of smoking

Significantly more participants with mental illness than controls smoked tobacco. Univariate ANOVA was used to examine the interaction between smoking status, metabolic and cardiovascular risk parameters. Each variable was entered into the model with group and smoking status as factors. With the exception of BMI (\( F = 4.25, \text{d.f.} = 1,93, P = 0.04 \)), there was no group x smoking status interaction.

### Effect of diagnosis

The impact of diagnostic group on metabolic and cardiovascular risk was examined. All metabolic and cardiovascular risk parameters were entered into a one-way ANOVA with diagnostic group (bipolar disorder, schizophrenia, schizoaffective disorder, other) as the factor in the model. There were no statistical differences in any of the variables between diagnostic groups.

### Effect of antipsychotic treatment

In order to investigate the interaction between the type of antipsychotic treatment (i.e. no treatment, atypical, typical or combination) and metabolic/cardiovascular risk parameters, all variables were entered into a one-way ANOVA with treatment group as the factor in the model. Serum insulin was significantly higher in participants taking atypical agents compared with all other groups (\( F = 2.8, \text{d.f.} = 3,173, P = 0.04 \)). There were no other statistically significant differences between treatment groups.

### Treatment of metabolic dysfunction and cardiovascular risk factors

The proportion of patients receiving appropriate pharmacological treatment for cardiovascular risk factors (hypertension and dyslipidaemia) was examined.
Main findings

The current study sought to investigate markers of metabolic dysfunction and cardiovascular risk estimates in a diagnostically heterogeneous sample of people with severe mental illness treated in the community. Compared with controls, people with mental illness, irrespective of diagnosis, had a significantly higher BMI (the mean BMI of 29.9 being within the overweight category and marginally short of the obese), waist circumference and waist-to-hip ratio (reflecting increased visceral adiposity). Dyslipidaemias and disorders of glucose homeostasis were more prevalent, as was the metabolic syndrome diagnosed according to the definition of the International Diabetes Federation (Alberti et al., 2006). The mean 10-year risk for cardiovascular disease (estimated according to both British and Framingham definitions) and the risk for a number of cardiovascular outcomes, including myocardial infarction and death due to cardiovascular disease, were consistently higher in participants with mental illness compared with controls. Moreover, a high proportion of people whose level of cardiovascular risk exceeds the threshold for intervention are not receiving appropriate treatment.

Other studies

Osborne et al. (2006) reported raised 10-year coronary heart disease risk scores (based on Framingham criteria), HDL cholesterol levels, total cholesterol level and an increased prevalence of diabetes mellitus in a sample of people with schizophrenia or non-affective psychoses from primary care. Another study has also reported increased 10-year cardiac risk in people with schizophrenia from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study (Goff et al., 2005). Correll et al. (2006) studied the prevalence of the metabolic syndrome and 10-year risk of coronary heart disease in psychiatric inpatients from across the diagnostic spectrum receiving atypical antipsychotics. Thirty-seven per cent of patients in this sample met National Cholesterol Education Program criteria for metabolic syndrome, and 47% fulfilled International Diabetes Federation criteria (Correll et al., 2006). This study lacked a control group, and although the prevalence of defined metabolic syndrome was higher than in our study, differences in participant characteristics (i.e. we studied community out-patients treated with typical and atypical antipsychotics), and a greater overall prevalence of obesity and the metabolic syndrome in

### Table 3  Metabolic parameters in participants with mental illness and controls

<table>
<thead>
<tr>
<th></th>
<th>People with mental illness (n=90)</th>
<th>Controls (n=92)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BMI, kg/m²: mean (s.d.)</strong>*</td>
<td>29.9 (4.9)</td>
<td>25.6 (4.6)</td>
</tr>
<tr>
<td><strong>Underweight, n (%)</strong></td>
<td>2 (2.2)</td>
<td>1 (1.1)</td>
</tr>
<tr>
<td><strong>Normal, n (%)</strong></td>
<td>9 (10)***</td>
<td>42 (45.7)</td>
</tr>
<tr>
<td><strong>Overweight, n (%)</strong></td>
<td>39 (43.3)</td>
<td>35 (38.0)</td>
</tr>
<tr>
<td><strong>Obese, n (%)</strong></td>
<td>40 (44.4)</td>
<td>14 (15.2)</td>
</tr>
<tr>
<td><strong>Blood pressure, mmHg: mean (s.d.)</strong>*</td>
<td>113.7 (15.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Systolic</strong></td>
<td>116 (19.9)</td>
<td>69.6 (9.9)</td>
</tr>
<tr>
<td><strong>Diastolic</strong></td>
<td>70 (11.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Waist circumference, cm: mean (s.d.)</strong>*</td>
<td>96.6 (13.1)***</td>
<td>84.1 (13.7)</td>
</tr>
<tr>
<td><strong>Waist-to-hip ratio: mean (s.d.)</strong>*</td>
<td>0.89 (0.09)***</td>
<td>0.82 (0.09)</td>
</tr>
<tr>
<td><strong>Total cholesterol</strong>*</td>
<td>5.7 (1.4)**</td>
<td>5.2 (0.9)</td>
</tr>
<tr>
<td><strong>HDL cholesterol</strong>*</td>
<td>1.3 (0.4)**</td>
<td>1.6 (0.4)</td>
</tr>
<tr>
<td><strong>LDL cholesterol</strong>*</td>
<td>3.4 (1.2)**</td>
<td>3.1 (0.9)</td>
</tr>
<tr>
<td><strong>Triglycerides</strong>*</td>
<td>2.1 (1.3)**</td>
<td>1.2 (0.6)</td>
</tr>
<tr>
<td><strong>Glucose homeostasis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fasting blood glucose, mmol/l: mean (s.d.)</strong>*</td>
<td>5.5 (1.4)*****</td>
<td>4.8 (0.5)</td>
</tr>
<tr>
<td><strong>HbA₁c, %: mean (s.d.)</strong></td>
<td>5.6 (0.9)***</td>
<td>5.2 (0.4)</td>
</tr>
<tr>
<td><strong>Serum insulin, mU/l: mean (s.d.)</strong>*</td>
<td>11.1 (8.1)*****</td>
<td>7.2 (5.1)</td>
</tr>
<tr>
<td><strong>HOMA-beta-cell function, %: mean (s.d.)</strong>*</td>
<td>98.8 (38.8)***</td>
<td>93.6 (36.9)</td>
</tr>
<tr>
<td><strong>HOMA insulin sensitivity, %: mean (s.d.)</strong>*</td>
<td>98.6 (55.6)*****</td>
<td>147.5 (72.0)</td>
</tr>
<tr>
<td><strong>HOMA insulin resistance, %: mean (s.d.)</strong>*</td>
<td>1.47 (1.1)*****</td>
<td>0.93 (0.64)</td>
</tr>
<tr>
<td><strong>Normoglycaemia, n (%)</strong></td>
<td>77 (85.6)</td>
<td>91 (98.9)</td>
</tr>
<tr>
<td><strong>Impaired fasting glucose, n (%)</strong></td>
<td>8 (8.9)**</td>
<td>1 (1.1)</td>
</tr>
<tr>
<td><strong>Diabetes mellitus, n (%)</strong></td>
<td>5 (5.5)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Metabolic syndrome, n (%)</strong></td>
<td>Yes</td>
<td>30 (33.3)*****</td>
</tr>
<tr>
<td><strong>No</strong></td>
<td>60 (66.7)</td>
<td>81 (88)</td>
</tr>
</tbody>
</table>

BMI, body mass index; HDL, high-density lipoprotein; LDL, low-density lipoprotein; HbA1c, glycated haemoglobin; HOMA, Homeostatic Model Assessment.

*P<0.05, **P<0.01, ***P<0.001.

**Dyslipidaemia**

Current recommendations state that treatment of dyslipidaemia should be based on an overall assessment of risk rather than an isolated serum lipid value. However, ‘high-risk’ patients should be offered prophylactic lipid-lowering therapy. Of the 13 high-risk patients, only 4 (30.8%) were receiving lipid-lowering therapy. One control participant was considered to be ‘high risk’ and was receiving appropriate therapy.

**Hypertension**

Hypertension was considered to be present if systolic blood pressure was ≥ 135 mmHg and/or diastolic blood pressure was ≥ 85 mmHg (Alberti et al., 2006). Fifteen participants with mental illness (16.7%) met criteria for hypertension compared with 13 controls (14.1%). Nine participants with mental illness and hypertension (60%) were not receiving an antihypertensive agent, compared with 10 controls with hypertension (77%).

**DISCUSSION**

Current research is increasingly adding to the weight of evidence that the burden of physical comorbidity is greater in people with severe mental illness. Studies investigating the prevalence of metabolic dysfunction and cardiovascular risk have focused largely on people with schizophrenia, and comparative data in other diagnostic groups are sparse.

**AUTHOR’S PROOF**

MACKIN ET AL

26
Our findings confirm the results of several other studies, and offer further insights into the nature of metabolic disease and cardiovascular disease risk in severe mental illness. The inclusion of a diagnostically heterogeneous sample is important in terms of understanding the effect of diagnosis on the development of metabolic dysfunction and cardiovascular risk. The problem of physical comorbidity and strategies for improving physical health in schizophrenia have been addressed previously (Marder et al., 2004). However, research in other psychiatric disorders such as bipolar disorders, lags behind (Clement et al., 2003) and there is an urgent need to establish whether there is a similar pattern of physical comorbidity. A high prevalence of metabolic syndrome and cardiovascular risk in psychiatric in-patients from across the diagnostic spectrum has recently been reported (Correll et al., 2006), and we confirm these findings in a sample of community out-patients treated with typical and atypical antipsychotics.

Investigating metabolic dysfunction in a community out-patient sample overcomes, to some extent, the confounding impact of physical inactivity on glucose homeostasis (Fulton-Kehoe et al., 2001) which is inherent in studies of psychiatric in-patients (Martinsen et al., 1989). All our participants were considered to be clinically stable, and thus the confounding influence on metabolic function of acute stress resulting from psychosis (Shiloah et al., 2003) or other distressing psychiatric symptoms was avoided.

Unlike previous studies investigating metabolic disease and cardiovascular risk, we also measured serum insulin and calculated insulin sensitivity and beta-cell function using the HOMA. Serum insulin and insulin resistance, both established independent risk factors for cardiovascular disease (Reaven, 2002), were increased in participants with mental illness. However, the mechanism underpinning the pathophysiology of insulin resistance in severe mental illness is poorly understood.

Although much of the current literature focuses on the risk of metabolic dysregulation in people taking atypical antipsychotics, significant numbers of people continue to take first-generation agents. Our study was designed to gather data on metabolic dysfunction and cardiovascular risk in a typical clinical setting. Although the study was not designed or powered to investigate the contribution of specific antipsychotic drugs, or classes of drugs, to metabolic or cardiovascular disease, with the exception of serum insulin levels, which were significantly higher in people taking atypical antipsychotics, the metabolic and cardiovascular risk profiles were similar in those taking typical, atypical or no antipsychotic medication. However, the small sample who were not receiving antipsychotic medication at the time of investigation had previously been prescribed an antipsychotic drug; any impact of this drug on metabolic function might have continued after the drug was no longer prescribed.

A further unique contribution of this study is the estimation of a number of cardiovascular outcomes. There is a striking and consistent difference in cardiovascular risk across a number of domains between people with mental illness and controls. Cardiovascular risk estimates were based on robust models derived from the JBS and the Framingham risk charts. These are frequently used by physicians to guide management of high-risk patients and to assist in decisions regarding intervention. Our data suggest that a high proportion of people with mental illness who are at high risk for adverse cardiovascular
events are not offered appropriate prophylactic intervention. This is in keeping with another recent study that has reported low rates of treatment for hypertension, dyslipidaemia and diabetes in people with schizophrenia from the CATIE trial (Nasrallah et al., 2006).

Limitations of the study

Although we did not detect differences in the prevalence of metabolic disease or estimates of cardiovascular risk across the diagnostic groups, the study might not have been sufficiently powered to detect such differences.

Selecting an appropriate control group for studies of this nature is complex. We attempted to control for demographic characteristics by specifically targeting carers and family members, and by recruiting controls from the geographical locale of participants with mental illness. This methodology might be considered somewhat crude, and as our analysis did not control for socio-economic variables we cannot exclude the possibility that the disparity in rates of metabolic disease and increased cardiovascular risk estimates are attributable to differing levels of deprivation.

People who volunteer to participate in medical research may take a more active interest in their physical health, and thus the prevalence of metabolic dysfunction and cardiovascular risk in the general population without severe mental illness might have been underestimated in our control group. The existence of such a potential bias is supported by the observed low prevalence of tobacco smoking in the control group (14%) compared with the reported prevalence in the general population. We cannot exclude the possibility, however, that a similar selection bias occurred in the recruitment of participants with mental illness: only 40% of this group smoked, which is lower than the prevalence (51%) reported in a recent large retrospective cohort study of people with severe mental illness (Osborn et al., 2007). These potential sources of bias may have resulted in an underestimate of the true prevalence of risk in both groups.

Although most of our participants with mental illness were taking antipsychotic medication at the time of investigation, the direction of causality cannot be established. There is accumulating evidence that antipsychotic drugs add to the metabolic burden in people with severe mental illness, but physical inactivity and diet are probably also influential. Tobacco smoking is also a well-established risk factor for cardiovascular disease (Unal et al., 2005), and although significantly more people with mental illness smoked compared with controls, differences in smoking behaviour did not account for the excess metabolic and cardiovascular risk. A genetic contribution to the increased metabolic and cardiovascular risk in people with severe mental illness should also be considered, as an increased prevalence of type 2 diabetes mellitus has been reported in unaffected first-degree relatives of people with schizophrenia (Mukherjee et al., 1989). This may suggest shared loci of genetic susceptibility for severe mental illness and diabetes, but shared environmental factors may also be important.

Fig. 1 Ten-year estimates for risk of adverse cardiovascular outcomes in people with mental illness (—) and controls (—). CVD, cardiovascular disease; CHD, coronary heart disease; JBS, Joint British Societies.
Implications

Current models of care appear to be failing a large proportion of people with severe mental illness. The reasons for this are likely to be manifold. Use of physical healthcare services often decreases after the onset of a psychiatric disorder (Jeste et al., 1996), and even when patients are engaged with healthcare services, rates of undiagnosed physical illnesses are often high (Mackin et al., 2005). Other factors may also contribute to poor detection and diagnosis of physical illness, including impaired ability to verbalise concerns (Lieberman & Coburn, 1986; Massad et al., 1990), poor insight into illness (Massad et al., 1990), depression (Goldman, 1999), or an unwillingness to consult a doctor other than their psychiatrist. When patients are cared for by psychiatrists, primary care physicians and physicians from other disciplines, there may be a shared assumption that a colleague is taking responsibility for managing a particular medical problem, when in fact the problem is not being attended to at all.

There are few studies specifically examining the impact of differing models of care on physical well-being and comorbidity in severe mental illness. One randomised trial from the USA evaluated an integrated model of primary medical care for a cohort of people with serious mental disorders, and the authors concluded that on-site, integrated primary care was associated with improved quality and outcomes of medical care (Druss et al., 2001). Interventions such as improving provider competencies through education and profiling, and organisational interventions such as computerised reminders to prompt mental health professionals to refer to primary care for appropriate screening, require further investigation.

There is a need for greater communication and collaboration between primary and secondary care, and for the establishment of clear guidelines outlining responsibilities and protocols for screening and managing physical health and disease in people with severe mental illness. Integrated models of care, including mental and physical health professionals, may be more appropriate for delivering care to this group.

Acknowledgements

The study was supported by a Research and Development Grant from the Newcastle, North Tyneside and Northumberland Mental Health NHS Trust. PM is a Department of Health Clinician Scientist Fellow.

PAUL MACKIN, PhD, MRCPsych, DAVID BISHOP, MRes, HELEN WATKINSON, BSc, PETER GALLAGHER, MPhil, I, NICOL FERRIER, MD, FRCPsych, School of Neurology, Neurobiology and Psychiatry, University of Newcastle upon Tyne, UK

Correspondence: Dr Paul Mackin, School of Neurology, Neurobiology and Psychiatry, University of Newcastle upon Tyne, Leazes Wing (Psychiatry), Royal Victoria Infirmary, Newcastle upon Tyne NE1 4LP, UK. Email: paul.mackin@ncl.ac.uk

(First received 2 October 2006, final revision 21 February 2007, accepted 21 March 2007)

REFERENCES


Rates and correlates of employment in people with schizophrenia in the UK, France and Germany

STEVEN MARWAHA, SONIA JOHNSON, PAUL BEBBINGTON, MAI STAFFORD, MATTHIAS C. ANGERMeyer, TRAOLACH BRUGHA, JEAN-MICHEL AZORIN, REINHOLD KILIAN, KARINA HANSEN and MONDHER TOUMI

Background Little is known about international variations in employment rates among people with schizophrenia or about the factors associated with employment in this disorder.

Aims To describe employment patterns and the variables associated with working in an international sample of people with schizophrenia.

Method An analysis was made of baseline data from the European Schizophrenia Cohort study, a 2-year investigation of people with schizophrenia in contact with secondary services and living in France, Germany and the UK (n = 1208).

Results Participants were working in all sections of the job market. People who had a degree, were living with their families or had experienced only a single episode of illness were more likely to be working. A continuous illness course, more severe non-psychotic symptoms and drug misuse reduced the odds of employment. There were large variations between centres in employment rates, which were highest in the three German and French centres by 8 to 35% (Gaite et al., 2002). Most previous studies have had small sample sizes, or recruited exclusively from rehabilitation services. The aim of our study was to examine employment patterns and variables associated with working in a large representative sample of people with schizophrenia resident in Germany, France and the UK. No previous analysis has been international and based on so large a sample.

METHOD

The European Schizophrenia Cohort (EuroSC) study is a naturalistic, 2-year follow-up of people aged 18–64 years who have schizophrenia and are in contact with secondary psychiatric services. A full explanation of the rationale and methods of the study, together with a description of the mental health services from which the samples were drawn, is presented by Bebbington et al. (2005).

Study sites

In France, participants were recruited from three centres: Lille (northern France), Lyon and Clermont-Ferrand (central France) and Marseille and Toulon (southern France). These centres are referred to hereafter as Lille, Lyon and Marseille. In former East Germany, participants were recruited from Leipzig and from the Altenburg area and the smaller towns and villages surrounding it. Data from these areas were pooled, and are referred to subsequently as the Leipzig centre. In West Germany, recruitment was from the Hemer district, including the cities of Hemer, Iserlohn and Werndorf, and from the Heilbronn district, including the city of Eppingen and surrounding towns and villages. The centres in the UK were the county of Leicestershire (excluding the city of Leicester but including surrounding towns and villages) and the inner-London borough of Islington.

Sampling

The participants were selected to provide a representative sample of people with schizophrenia under the care of secondary mental health services. People who had recently lost contact with services were not included, except in the London sample. Sampling was achieved in London and in all the French centres by establishing a list of all people with a psychotic diagnosis in the areas from information already kept by the mental health services and taking a random sample from all those identified. In the London centre patients were randomly sampled from the whole local list, whereas in France sampling was stratified so that within each centre ten patients were randomly selected from each of the ten local sectors.

In Germany and in Leicestershire, lists of all potential participants in each catchment area were compiled and all eligible people were included in the sample. In all centres, a diagnosis of schizophrenia was confirmed after an interview using structured instruments by a study investigator who applied the DSM–IV criteria (American Psychiatric Association, 1994). Patients were eligible for inclusion if they were aged 18–64 years, had a diagnosis of schizophrenia according to DSM–IV criteria and gave informed consent. Patients who had been continuously in hospital for the previous 12 months, or were currently homeless or planning to move (and therefore unavailable for follow-up) were excluded.

Instruments

An extensive battery of instruments was used to collect information at interview, but only those relevant to this analysis are presented here. Employment data were obtained through the Lehman Quality of Life Interview (Lehman, 1983); this establishes...
whether participants are currently employed and, if relevant, their job title. Questions on receipt of welfare benefits are also included in this interview. Psychiatric and social history, including educational history and whether participants had ever been employed, was recorded using the Past History and Socio-demographic Description Schedule (World Health Organization, 1973).

The Schedule for Clinical Assessment in Neuropsychiatry (SCAN; Wing et al, 1990; World Health Organization, 1992) was used to establish a DSM-IV diagnosis of schizophrenia in the UK and Germany, whereas the French centres used the Structured Clinical Interview for DSM-IV (First et al, 1997). Part of the alcohol and substance misuse data available from the SCAN allowed the coding of variables indicating lifetime history of alcohol and drug misuse. Information on the participants’ current symptom profile was collected through the Positive and Negative Syndrome Scale (PANSS; Kay et al, 1987, 1989). The summary indicators from the PANSS used in this analysis were the positive, negative and general psychopathology symptom sub-scores. The general psychopathology section covers a wide variety of non-psychotic symptoms, and includes measures of anxiety and depression, uncooperativeness, cognitive symptoms, impulse control and social avoidance.

Adherence was evaluated using the Rating of Medication Influences scale (ROMI; Weiden et al, 1994). This assesses subjective attitudes and behaviours that influence patients’ compliance with antipsychotic medication. It yields a total ‘reasons for compliance’ score, higher in people who are more willing to take medication, and a ‘reasons for non-compliance’ score, higher in people who are more reluctant.

The European Service Mapping Schedule (ESMS; Johnson et al, 2000) was used to describe and classify the mental health services available in all three sites. Vocational services are known to affect the employment rate in schizophrenia (Drake et al, 1998), and this study allowed us to consider if any variation in employment rate between sites might be linked to employment service provision.

Combined training was held involving interviewers from all three sites, to try to ensure instruments were used reliably. However, no formal assessment was made of reliability between centres. For those instruments not already available in French and German, translation followed the comprehensive procedures recommended by the World Health Organization (Sartorius & Kuyken, 1994), including back-translation.

Analysis
We analysed the baseline data of this 2-year cohort with the aims of describing patterns of employment and exploring which candidate explanatory variables were independently associated with likelihood of being employed in this large international sample. Descriptive analyses and univariate tests were carried out using the Statistical Package for the Social Sciences version 11.5 and the subsequent logistic regression using Stata version 8 for Windows. Employment was broadly defined as having any job, whether full-time or part-time and whether obtained through the open labour market, sheltered vocational schemes or the voluntary sector. A broad definition of working was used because quantitative and qualitative studies exploring the attitudes of people with schizophrenia to employment suggest that people desire a range of different types of employment, not only full-time competitive paid work (Secker et al, 2001; Honey, 2004; Marwaha & Johnson, 2005). Whether a person was working according to this definition was the main variable used in the analyses. Within the employed group we also examined how many in each centre were supporting themselves solely with earnings, and how many were working but also claiming some form of welfare benefits.

The analysis was conducted in a number of stages. Descriptive analyses were conducted of patterns of employment in each centre: current rates of employment and types of jobs; the number of people employed and supporting themselves through earnings only; and the percentage of people never employed.

Using χ² tests for categorical outcomes and t-tests for continuous outcomes, we analysed the association of employment with various explanatory variables derived from a literature review on correlates of employment among people with schizophrenia (Marwaha & Johnson, 2004). These correlates were educational history, negative and positive symptom severity, gender, marital status, accommodation and living conditions. We also identified and tested the association with working of a number of other potential variables of interest: area of residence (study centre), general population employment rate in area of residence, adherence to medication, course of illness, age at illness onset, length of illness, lifetime history of alcohol or drug misuse and ethnic group.

Previous analyses of this data-set (Bebbington et al, 2005) had established the presence of differences between the three countries in social and clinical sample characteristics, and we therefore repeated each of these univariate analyses stratified by country. German centres did not code ethnicity in the same way as the UK and French sites, and German participants could therefore only be classified as ‘born in Germany’ or ‘German resident but born outside’. Separate analyses were therefore made for each country examining the association between ethnic group (or, in the case of Germany, where born) and employment. Logistic regression was then used to identify explanatory variables independently associated with employment using the ‘enter’ method, in which all independent variables are entered into the equation at the same time. Variables associated with employment on univariate analyses at the P < 0.1 level of significance were entered into the model as independent variables. Ethnicity was excluded as a variable in the logistic regression because of a lack of uniformity in the way it was categorised across the three countries. Lack of independence for observations regarding individuals within the same centre was allowed for in this analysis by computing robust standard errors, clustered on centre (Rogers, 1993).

RESULTS
In total, 1208 people with schizophrenia participated in the study: 288 in France, 302 in the UK and 618 in Germany. Gender distribution varied between countries: 64.6% of participants in the UK, 69.4% in France and 56.3% in Germany were male. The mean age was approximately 40 years in all three countries, but marital status and living conditions differed significantly. Similar numbers of participants were living alone in each country, but more German respondents were living with partners and/or children, and more French respondents with their parents. Participants in France were the most symptomatic, with a mean total PANSS score of 71, followed by scores of 56 in Germany and 48 in the UK. The clinical and socio-demographic
profiles of the sample are described in more detail by Bebbington et al (2005).

**Employment**

Table 1 shows employment rates in each centre. The employment rate in the general population in the regions in which the centres lie is also shown. The overall employment rate of participants was 21.5%, but varied between countries and sites, with rates of 12.9% in the UK, 11.5% in France and 30.3% in Germany. This compares with general population employment rates of 71.0% in the UK, 62.2% in France and 65.4% in Germany in the year 2000 (European Commission, 2002).

The proportion of people in each country who were supporting themselves entirely through working and were not receiving welfare benefits was 8.9% in the UK, 7.6% in France and 11.8% in Germany. The German centres thus had the highest proportion of people working as well as the highest proportion supporting themselves entirely through work, although the difference in the latter was less striking.

The number of people in each centre who had never been employed was low, apart from in Marseille.

Table 2 describes the jobs of study participants, using the UK Standard Occupational Classification 2000 (Office for National Statistics, 2000) for the sake of uniformity. In all three countries people with schizophrenia appeared to work in nearly all sections of the job market. The most common types of work were ‘elementary’ jobs, such as cleaning and labouring, and ‘skilled trade occupations’, such as plumbing or metalwork. The proportion of people working in senior official or managerial positions or as process plant and machine operatives was very small. More people in Germany were doing sheltered or voluntary work.

**Employment service characteristics of the three sites**

The main difference in employment service configurations, mapped using the ESMS, was that the German centres had more vocational services and more placements provided within these than the other two countries. This was particularly the case for high-intensity work services and high-intensity work-related services. The ESMS defines the former as services offering work placements 2 days a week but at a rate of pay below 50% of the local minimum wage. Of the German centres, Heilbronn had the greatest number of patients currently in placements, but mostly in supported work activities paid below 50% of the minimum wage. Few vocational services were recorded in France, but the French centres did not include social services or voluntary sector provision in their data collection. In London there were some work activities within day centre settings, but few formal sheltered or supported work schemes.

**Unadjusted analyses**

Employment was significantly associated with area of residence, having a diploma or degree, living conditions, alcohol misuse, general population employment rate in area

<table>
<thead>
<tr>
<th>Occupation¹</th>
<th>UK n (%)²</th>
<th>France n (%)</th>
<th>Germany n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managers and senior officials</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Professional occupations</td>
<td>4 (10)</td>
<td>1 (3)</td>
<td>9 (5)</td>
</tr>
<tr>
<td>Associate professional and technical occupations</td>
<td>5 (13)</td>
<td>1 (3)</td>
<td>7 (4)</td>
</tr>
<tr>
<td>Administrative and secretarial occupations</td>
<td>3 (8)</td>
<td>2 (6)</td>
<td>6 (3)</td>
</tr>
<tr>
<td>Skilled trades occupations</td>
<td>7 (18)</td>
<td>5 (15)</td>
<td>14 (8)</td>
</tr>
<tr>
<td>Personal service occupations</td>
<td>3 (8)</td>
<td>2 (6)</td>
<td>7 (4)</td>
</tr>
<tr>
<td>Sales and customer service occupations</td>
<td>3 (8)</td>
<td>1 (3)</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Process, plant and machine operatives</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Elementary occupations</td>
<td>11 (28)</td>
<td>8 (24)</td>
<td>41 (22)</td>
</tr>
<tr>
<td>Voluntary/sheltered work</td>
<td>2 (5)</td>
<td>6 (18)</td>
<td>71 (38)</td>
</tr>
<tr>
<td>Difficult to classify</td>
<td>0 (0)</td>
<td>4 (12)</td>
<td>15 (8)</td>
</tr>
<tr>
<td>Job data missing</td>
<td>0 (0)</td>
<td>2 (6)</td>
<td>11 (6)</td>
</tr>
<tr>
<td>Total number working</td>
<td>39</td>
<td>33</td>
<td>187</td>
</tr>
</tbody>
</table>

1. Classified according to the Standard Occupation Classification 2000 UK.
2. Percentages of the total group who are working within each country.
Table 3  Work and social variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number working/ total in group</th>
<th>Percentage of group working</th>
<th>Association with working</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n/N</td>
<td>%</td>
<td>( \chi^2 )</td>
</tr>
<tr>
<td>Vocational training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>99/524</td>
<td>18.9</td>
<td>3.16</td>
</tr>
<tr>
<td>Yes</td>
<td>147/634</td>
<td>23.2</td>
<td></td>
</tr>
<tr>
<td>Diploma or degree</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>210/1026</td>
<td>20.5</td>
<td>6.82</td>
</tr>
<tr>
<td>Yes</td>
<td>43/143</td>
<td>30.1</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>162/742</td>
<td>21.8</td>
<td>1.23</td>
</tr>
<tr>
<td>Female</td>
<td>96/459</td>
<td>20.9</td>
<td>0.15</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single, divorced or widowed</td>
<td>167/743</td>
<td>22.5</td>
<td></td>
</tr>
<tr>
<td>Married or cohabiting</td>
<td>61/254</td>
<td>24.0</td>
<td></td>
</tr>
<tr>
<td>Living conditions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone at home</td>
<td>77/417</td>
<td>18.5</td>
<td></td>
</tr>
<tr>
<td>With partner (with or without children)</td>
<td>62/268</td>
<td>23.1</td>
<td></td>
</tr>
<tr>
<td>With family (not partner)</td>
<td>69/293</td>
<td>23.5</td>
<td>15.60</td>
</tr>
<tr>
<td>Supported housing</td>
<td>42/141</td>
<td>29.8</td>
<td></td>
</tr>
<tr>
<td>Homeless and other</td>
<td>9/88</td>
<td>10.3</td>
<td></td>
</tr>
<tr>
<td>Accommodation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owner-occupied</td>
<td>40/176</td>
<td>22.7</td>
<td></td>
</tr>
<tr>
<td>Rented</td>
<td>161/799</td>
<td>20.2</td>
<td>Fisher’s 0.347</td>
</tr>
<tr>
<td>Supported accommodation</td>
<td>42/155</td>
<td>27.1</td>
<td>Fisher’s exact 0.604</td>
</tr>
<tr>
<td>Hospital</td>
<td>12/65</td>
<td>18.5</td>
<td>Test 0.038</td>
</tr>
<tr>
<td>Homeless</td>
<td>0/2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>5/10</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>Substance use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol misuse</td>
<td>35/223</td>
<td>15.7</td>
<td>5.68</td>
</tr>
<tr>
<td>No alcohol misuse</td>
<td>224/975</td>
<td>23.0</td>
<td></td>
</tr>
<tr>
<td>Drug misuse</td>
<td>13/90</td>
<td>14.4</td>
<td>2.08</td>
</tr>
<tr>
<td>No drug misuse</td>
<td>243/1104</td>
<td>22.0</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>33/234</td>
<td>14.1</td>
<td>Fisher’s 0.804</td>
</tr>
<tr>
<td>Black Caribbean</td>
<td>4/42</td>
<td>9.5</td>
<td>Exact 0.944</td>
</tr>
<tr>
<td>Asian</td>
<td>2/13</td>
<td>15.4</td>
<td>Test 0.704</td>
</tr>
<tr>
<td>Turkish/Greek</td>
<td>0/6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0/7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>30/269</td>
<td>11.2</td>
<td>Fisher’s 0.347</td>
</tr>
<tr>
<td>Black Caribbean/African</td>
<td>1/3</td>
<td>33.3</td>
<td>Exact 0.68</td>
</tr>
<tr>
<td>Turkish/Greek</td>
<td>0/1</td>
<td>0</td>
<td>Test 0.001</td>
</tr>
<tr>
<td>Other</td>
<td>2/14</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Born in Germany</td>
<td>168/577</td>
<td>29.1</td>
<td>5.63</td>
</tr>
<tr>
<td>German but born abroad</td>
<td>17/34</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>Local employment rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51–55%</td>
<td>8/101</td>
<td>7.9</td>
<td>Test for (&lt;0.0005)</td>
</tr>
<tr>
<td>56–60%</td>
<td>15/87</td>
<td>17.2</td>
<td>Linear 0.001</td>
</tr>
<tr>
<td>61–65%</td>
<td>147/767</td>
<td>19.2</td>
<td>Trend 0.001</td>
</tr>
<tr>
<td>66–70%</td>
<td>60/100</td>
<td>60.0</td>
<td>19.4</td>
</tr>
<tr>
<td>71–75%</td>
<td>29/152</td>
<td>19.1</td>
<td></td>
</tr>
</tbody>
</table>

Years of education: workers mean 10.2 (s.d. = 2.0); non-workers, mean 10.2 (s.d. = 2.1); \( t = 0.62, p = 0.54 \)

of residence, more severe positive, negative and general psychopathology symptoms, longer length of illness and illness course (Table 3–5). In Germany, foreign-born people were more likely to be working. The associations with illness variables were particularly highly significant on these analyses, as were the associations with area of residence. On repeating each of the analyses in Tables 3–5 stratifying by country, the direction of the relationships remained essentially the same, although reduced power rendered some of the associations non-significant at the 5% level.

Logistic regression

Table 6 shows the final regression model, which explained 19% of the variance in employment status. Vocational training, regional employment rate, negative symptom score, alcohol misuse, duration of illness and reasons for non-compliance score were all entered into the regression, but did not reach the \( P = 0.05 \) threshold for significance and were omitted from the final model.

People with schizophrenia living in Leicestershire, Marseille, Leipzig, Hemer and Heilbronn all had higher odds of being employed than those living in London. Living with family (other than a partner), having a degree or a diploma, and having experienced only one episode of illness with full remission also improved the odds of working. A continuous illness course, higher general psychopathology scores on the PANSS, earlier onset of illness and a history of drug misuse all reduced the odds of current employment. A higher PANSS positive symptom score was associated with a greater likelihood of working, a reversal of the effect observed on unadjusted analysis, but this effect was a weak one.

**DISCUSSION**

Comparison of employment rates between centres

This is the first international comparative study to report on employment patterns and on variables associated with working in people with schizophrenia. As in other recent studies, the employment rate in the UK sample was low, especially in London (UK700 Group, 1999; Perkins & Rinaldi, 2002). The rate in France is equally concerning, at approximately a third of the German rate. The large numbers of unemployed people with schizophrenia represent a significant financial cost (Huxley &
Table 4 Work and clinical variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Workers (n=259) mean (s.d.)</th>
<th>Those not working (n=948) mean (s.d.)</th>
<th>Difference between workers and non-workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>PANSS score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive symptoms</td>
<td>11.2 (4.9)</td>
<td>12.7 (5.7)</td>
<td>4.22 &lt; 0.0005</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>14.1 (6.9)</td>
<td>16.2 (7.8)</td>
<td>4.34 &lt; 0.0005</td>
</tr>
<tr>
<td>General psychopathology</td>
<td>26.5 (9.4)</td>
<td>30.0 (10.8)</td>
<td>5.13 &lt; 0.0005</td>
</tr>
<tr>
<td>Length of illness, years</td>
<td>14.3 (9.1)</td>
<td>16.0 (10.0)</td>
<td>2.61 0.009</td>
</tr>
<tr>
<td>Age at illness onset, years</td>
<td>25.5 (7.7)</td>
<td>26.5 (8.5)</td>
<td>1.70 0.090</td>
</tr>
<tr>
<td>ROMI score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reasons for compliance</td>
<td>2.14 (0.44)</td>
<td>2.09 (0.45)</td>
<td>1.37 0.17</td>
</tr>
<tr>
<td>Reasons for non-compliance</td>
<td>1.68 (0.61)</td>
<td>1.60 (0.56)</td>
<td>−1.78 0.075</td>
</tr>
</tbody>
</table>

Table 5 Work and illness course

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number working/ total in group n/N</th>
<th>Percentage working in category %</th>
<th>χ²</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall illness course</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Episodic with inter-episode symptoms</td>
<td>123/509</td>
<td>24.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Episodic without inter-episode symptoms</td>
<td>75/234</td>
<td>32.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous</td>
<td>25/304</td>
<td>8.2</td>
<td>52.07 &lt; 0.0005</td>
<td></td>
</tr>
<tr>
<td>Single episode in partial remission</td>
<td>10/45</td>
<td>22.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single episode in full remission</td>
<td>14/46</td>
<td>30.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other/unspecified course</td>
<td>9/54</td>
<td>16.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illness course characterised by prominent negative symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>76/395</td>
<td>19.2</td>
<td>1.67 0.20</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>182/809</td>
<td>22.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Those not working in:

1.68 (0.61) 1.60 (0.56) −1.78 0.075

PANSS, Positive and Negative Syndrome Scale; ROMI, Rating of Medication Influences.

1. There are missing values for some variables: the greatest number of missing values is for reasons for compliance, where there are 935 ratings for the non-workers and 256 for the workers.

Leicestershire had a higher employment rate. City living also appears to have an adverse effect on employment in people with severe mental illness in the USA (Drake et al, 1998). The large variations between centres, which were not accounted for by differences in clinical and socio-demographic characteristics of the samples, suggest that social and service factors that vary between countries and regions are likely to be important.

Vocational services, benefits payments and employment rates

The variations between areas in employment rates were even larger than those between countries, and persisted when adjustment was made for patient characteristics and regional general population employment through regression analysis. Various societal factors that might account for these variations have already been discussed. Another possible explanation is the substantially greater provision of...
vocational services and placements in Germany. German centres had both the
largest numbers working in sheltered set-
tings and relatively large numbers support-
ing themselves entirely through their earn-
ings without recourse to state benefits. This
suggests that the good availability of
opportunities for combining work and
benefits does not necessarily lead to people
getting stuck in this situation, and may be
associated with a greater proportion of
people supporting themselves fully through
open market earnings, although we cannot
say whether there is a causal relationship.

**Occupation**

There was no large difference between the
proportions of people in each country who
worked in various types of job, but the
numbers were small. Previous studies have
suggested that people with schizo-
phrenia tend to be in positions involving
low levels of interpersonal interaction
(Morgan & Gopalswamy 1983; Bacani-
Oropilla et al, 1991). ‘Elementary’ occupa-
tions such as cleaning and labouring
may be of this nature, as well as some
skilled trades. Both were relatively highly
represented.

Few people with schizophrenia were in
managerial or senior official positions
where training periods may be long, a series
of promotions required and an episodic
illness thus particularly damaging. It is
surprising that more people were not
working in information technology, given
the expansion of this industry over the past
decade. Overall, it appears that a diagnosis
of schizophrenia is probably not a bar to
doing any kind of job, but makes entry into
certain job types less likely.

**Social correlates of employment
in people with schizophrenia**

Living with family (other than a partner)
was associated with a greater likelihood of
working, persisting after adjusting for other
independent variables. This may reflect
better social support, enabling better social
recovery. Ethnic group was not significa-
cantly associated with employment in the
UK and French samples, although place of
birth did emerge as important in Germany
in the univariate analyses. Small numbers
of minority ethnic group members limit
conclusions about this. We did not explore
in any detail the effects of cultural back-
grounds and values on whether people
worked, but these may have a role in
explaining variations in employment rates.

Having a diploma or degree was signifi-
cantly associated with employment. Acqui-
sition of a tertiary education may reflect
generally better premorbid functioning,
associated with better overall outcomes, and
possessions of some qualifications may also
broaden choice for those seeking to re-enter
the labour market.

**Clinical correlates of employment**

Our finding that comorbid drug misuse was
associated with lower odds of working sup-
ports a previous study indicating high levels
of social exclusion among people with ‘dual
diagnosis’ (Todd et al, 2004). Substance
misuse may well make job-seeking and
good occupational functioning more diffi-
cult. Earlier onset of illness also reduced
the odds of working, reflecting a poorer
overall prognosis in people who become
unwell earlier, and may also be related to
fewer opportunities for training and em-
ployment before the onset of illness.

There have been contradictory findings
in previous studies about whether the posi-
tive and negative symptoms of schizo-
phrenia are significantly associated with
employment status (Anthony & Jansen,
1984; Cook & Razzano, 2000). Positive
and negative symptoms were significantly
associated with employment in the unad-
justed analyses but no clear relationship
was found on multivariate analysis. There
was no significant association with negative
symptoms, but a marginally significant
\(P=0.038; OR=1.03, 95\% \text{ CI } 1–1.06\) ten-
dency for more positive symptoms to be
associated with a greater likelihood of
working. The lack of the anticipated re-
lationship between greater symptom sever-
ity and lower likelihood of working may
well be due to the association between
higher symptom scores and a continuous
illness course, which substantially reduced
the odds of employment. However, the pre-
sence of positive or negative symptoms
should not be thought of as necessarily pre-
venting people with schizophrenia return-
ing to the labour market.

We found that the symptoms assessed by
the general psychopathology section of the
PANSS were related to employment
status, although the effect was not large.
Psychiatrists may not routinely treat

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds ratio</th>
<th>(95% \text{ CI})</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Living in(^1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leicestershire</td>
<td>3.68 (2.84–4.77)</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Marseille</td>
<td>6.63 (2.02–21.78)</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Leipzig</td>
<td>3.70 (1.96–7.00)</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Hemer</td>
<td>9.54 (4.05–22.44)</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Heilbronn</td>
<td>36.37 (17.45–75.78)</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Living with family (not partner)(^2)</td>
<td>1.41 (1.12–1.78)</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>Diploma or degree</td>
<td>1.75 (1.07–2.88)</td>
<td>0.026</td>
<td></td>
</tr>
<tr>
<td>Drug misuse</td>
<td>0.27 (0.13–0.58)</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Illness course(^1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous illness course</td>
<td>0.44 (0.29–0.68)</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Single episode of illness in full remission</td>
<td>1.40 (1.24–1.57)</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Age at illness onset</td>
<td>0.98 per year (0.96–1.00)</td>
<td>0.033</td>
<td></td>
</tr>
<tr>
<td>General psychopathology symptoms (total score)</td>
<td>0.95 per point on PANSS sub-scale (0.93–0.98)</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Positive psychotic symptoms (total score)</td>
<td>1.03 per point on PANSS sub-scale (1.00–1.06)</td>
<td>0.038</td>
<td></td>
</tr>
</tbody>
</table>

PANSS, Positive and Negative Syndrome Scale
1. The comparison centre was London; Lille and Lyon were not significantly different from London.
2. Comparison category is ‘living alone’.
3. Comparison category is single episode, partial remission.
symptoms such as anxiety, depression, social avoidance, poor impulse control and cognitive deficits very vigorously, or treatment response may be poor. The presence of these symptoms may be seriously disabling.

In any cross-sectional analysis such as this, interpretation must be tentative, given that causation cannot be assigned. Prospective data are needed to answer the question of causation, and will be available from this data-set.

Limitations
Our definition of employment was a broad one, and we were unable to identify accurately the group who were in open-market employment, strictly defined, as full details of how and with what support people had obtained work were not collected. We did, however, measure how many people were working and supporting themselves completely through their pay and how many were working as well as claiming benefits.

Because the study did not formally check the reliability of ratings, country-level comparisons of data from interviewer-rated instruments such as the PANSS require cautious interpretation. However, employment status is less likely to be subject to problems of reliability. Although efforts were made to ensure consistent and comparable procedures in all centres, the service structures were different, and recruitment bias cannot be excluded. It is unlikely, however, to explain a large amount of the variation in employment rates, given that the UK sample who had the least severe symptoms in comparison with other countries had a similarly low rate of employment to the French sample, who scored highly on symptom severity.

Ideally, a hierarchical form of analysis, such as multilevel modelling, should be used to explore the effects of country and local level variables on employment status at individual level. However, although our study had excellent power for investigation of individual-level explanatory variables, the three countries and eight centres were insufficient for such a multilevel analysis. Our data therefore did not allow a substantial exploration of the effects of national and regional level variables on employment among the mentally ill population. In addition, attribution of regional employment rates to individuals is a somewhat unsatisfactory method of exploring the relationship between local and individual characteristics. In future investigations, inclusion of a larger number of centres, detailed investigations of their characteristics and use of multilevel modelling techniques would be desirable.

Regression modelling using sets of variables that have substantial intercorrelations, as in this study, results in models that are relatively susceptible to change with small alterations in the variables included. This should be borne in mind, especially in relation to variables that are marginally significant in the final model. A further caveat is that we did not use the Bonferroni correction to adjust for multiple testing. Thus findings that are close to the P=0.05 level of significance should be treated with caution, although most of the significant associations we found in the final regression were at least at the P<0.001 level.

Data on job history, which is an important predictor of future employment, were not collected. We are not able to provide information on non-responders.

ACKNOWLEDGEMENTS
Our thanks to all the patients and staff who helped with the study and to the Camden and Islington Mental Health and Social Care NHS Trust and the Leicestershire Partnership NHS Trust Research and Development Programme.


REFERENCES


Determining when impairment constitutes incapacity for informed consent in schizophrenia research

SCOTT Y. H. KIM, PAUL S. APPELBAUM, JEFFREY SWAN, T. SCOTT STROUP, JOSEPH P. MEEOVY, DONALD C. GOFF, DILIP V. JESTE, J. STEVEN LAMBERTI, ADRIAN LEIBOVICI and ERIC D. CAINE

Background Although people with schizophrenia display impaired abilities for consent, it is not known how much impairment constitutes incapacity.

Aims To assess a method for determining the categorical capacity status of potential participants in schizophrenia research.

Method Expert-judgement validation of capacity thresholds on the sub-scales of the MacArthur Competence Assessment Tool – Clinical Research (MacCAT–CR) was evaluated using receiver operating characteristic (ROC) analysis in 91 people with severe mental illness and 40 controls.

Results The ROC areas under the curve for the understanding, appreciation and reasoning sub-scales of the MacCAT–CR were 0.94 (95% CI 0.88–0.99), 0.85 (95% CI 0.76–0.94), and 0.80 (95% CI 0.70–0.90). These findings yielded negative and positive predictive values of incapacity that can guide the practice of investigators and research ethics committees.

Conclusions By performing such validation studies for a few categories of research with varying risks and benefits, it might be possible to create evidence-based capacity determination guidelines for most schizophrenia research.

Declaration of interest D.C.G. is on the advisory boards of several pharmaceutical companies. Funding detailed in Acknowledgements.

The ethics of research involving adults with impaired decision-making capacity remains a focus of policy discussions in the USA (Kim et al., 2004), of policy statements internationally (UNESCO, 2005), and a subject of new legislation in the UK (Adults with Incapacity (Scotland) Act 2000; Mental Capacity Act 2005) and two US states (Kim et al., 2004). In particular, research involving people with schizophrenia has been controversial because as a group they have greater decisional impairment than healthy controls (Carpenter et al., 2000; Kovalnick et al., 2003; Palmer et al., 2004). However, diagnosis cannot be equated with decisional incapacity because there is too much heterogeneity in decisional abilities (Grasso & Appelbaum, 1995b; Carpenter et al., 2000; Palmer et al., 2004). Although there are now instruments for assessing decisional ability, we currently lack an evidence-based method for translating those dimensional data into categorical judgements (Kim, 2006).

In this study, we used the judgements of independent clinicians experienced in capacity assessments to address the following question: given that people with schizophrenia exhibit a range of decisional abilities, how can we use a standardised instrument to distinguish those who are capable from those who are incapable of informed consent? We asked the question in the context of a unique opportunity presented by a multisite clinical trial, funded by the National Institute of Mental Health, the Clinical Antipsychotic Trials of Intervention Effectiveness – Schizophrenia (CATIE; Stroup et al., 2003), which used as part of its research protocol the most widely tested measure of decisional ability, the MacArthur Competence Assessment Tool – Clinical Research (MacCAT–CR; Appelbaum & Grasso, 2001).

METHOD

Participants
In line with the aim of the project, the goal of recruitment was to ensure that a sufficient spectrum of decision-making abilities was represented in our sample, rather than a random sample of a particular population. Participants included 91 people with severe mental illness and 40 people in the community comparison group. The group with severe mental illness consisted of two subgroups: 55 participants in the CATIE–Schizophrenia study at six different sites across the USA; and 36 people who were not part of the CATIE study but were recruited specifically for this interview study from two out-patient clinics serving people with severe and persistent mental illnesses, and from in-patient units at a state hospital in Rochester, New York, USA. Those who were not part of the CATIE study were added to ensure a sufficient spectrum (i.e. to avoid spectrum bias; Zhou et al., 2002) of decision-making ability; we noticed in the early part of the study that the performance of those in the CATIE study tended to cluster in the upper end – a trend that was ultimately borne out in the overall CATIE–Schizophrenia sample (Stroup et al., 2005). The participants in the control group were all without psychosis and were recruited in Rochester through advertisements in the community, in support staff work areas of a general hospital and at an out-patient substance misuse recovery programme.

This study was approved by the research ethics committees (institutional review boards) of all participating institutions, and all participants provided written informed consent after full disclosure of study elements. The CATIE participants provided separate informed consent for this ancillary study. For the group with severe mental illness, as has been done in other studies of this kind (Moser et al., 2002; Stroup et al., 2006), given the low risk of this interview study, a relatively undemanding standard for capacity to consent was used.

Measures
Participants were videotaped during their assessment with the MacCAT–CR (Appelbaum & Grasso, 2001). The MacCAT–CR has been extensively used in people with schizophrenia (Carpenter et al., 2000; Dunn et al., 2002; Moser et al., 2002; Stroup et al., 2005) and people with major depression (Appelbaum et al., 1999) and dementia (Kim et al., 2001), and is a companion instrument to the MacArthur Competence Assessment Tool for Treatment (MacCAT–T) (Cairns et al., 2005a,b).
The MacCAT–CR contains pertinent disclosure elements of informed consent and is designed to be adapted to specific research protocols, to reflect the task-specific nature of decisional capacity (Appelbaum & Grisso, 2001). The version used in the CATIE–Schizophrenia study was used for all participants in this study; thus, the non-CATIE and control participants were asked to imagine being invited to participate in the CATIE study as their decisional abilities were assessed. This procedure is commonly employed in capacity research (Carpenter et al., 2000; Moser et al., 2002).

The MacCAT–CR is structured according to the four-abilities model of decision-making capacity (Grisso & Appelbaum, 1998). These include ‘understanding’ [emphasis added] of disclosed information about the nature of the research project and its procedures (13 items for a possible total score of 26 – each item in the MacCAT–CR has a score range of 0–2 with objective scoring criteria); appreciation of the effects of research participation (or failure to participate) on subjects’ own situations (3 items for a possible total score of 6); reasoning about participation (4 items for a possible total score of 8); and ability to communicate a choice (one item for a possible total score of 2)’ (Appelbaum et al., 1999). Data on the ability to communicate a choice will not be discussed here as almost everyone received a full score. The MacCAT–CR does not provide a global score because requirements for each ability related to capacity can vary by jurisdiction and according to the decisional demands of a given study (Grisso & Appelbaum, 1995a). However, it is important to note that the four-abilities model is based on an extensive review of laws, court decisions and ethics literature, such that it provides a reasonable approximation of the standards for capacity broadly laid out in statutes. Thus, researchers have been able to use the MacCAT instruments to approximate, for example, the criteria of the Mental Capacity Act 2005 (Cairns et al., 2005a,b).

The final MacCAT–CR sub-scale ratings for all 131 participants used for analysis were made by J.S. During the course of the project, the principal investigator (S.K.) independently scored 36 out of the 131 interviews. This was the basis for calculations of interrater reliability. For those 36 participants, discrepancies arising after independent scoring of MacCAT–CR items by the two raters were resolved through discussion between the two raters. The intraclass correlation coefficients for total scores of MacCAT–CR sub-scales were 0.93 for understanding (F = 29.3, d.f. = 35.0, 35, P < 0.0001), 0.89 for appreciation (F = 16.7, d.f. = 35.0, 35; P < 0.0001), and 0.84 for reasoning (F = 11.3, d.f. = 35.0, 35, P < 0.0001).

Psychiatric diagnoses were made using medical records and the Structured Clinical Interview for DSM–IV (SCID–IV; First et al., 1997). Severity of psychiatric symptoms was measured using the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987), which includes positive, negative and general psychopathology sub-scales. Control participants were administered the SCID only.

**Expert judgements**

Three psychiatrists with experience in assessing decisional capacity (two consultation psychiatrists and one board-certified geriatric psychiatrist) were recruited to serve as expert judges and a fourth judge was added as a back-up. The judges were prepared for their task by informing them of the basic outlines of the CATIE–Schizophrenia study (the rationale, the medications to be tested, the total number of participants to be enrolled, and the fact that treatment failures would lead to randomisation with a new study drug). They were told that their job was to render a categorical judgement based on viewing an interview of a semi-structured capacity assessment (but they were unaware of the actual MacCAT–CR scores). The ultimate goal of deriving a final judgement was explained as: ‘Your task is to review the tapes carefully and make a categorical judgement (definitely capable, probably capable, probably not capable, and definitely not capable). In the real world, decisions need to be made even if things aren’t clear, as we reduce complex clinical data into a yes/no judgement’. They also rated the statement: ‘The videotaped interview gave me sufficient basis to make my decision in this case’ on a 5-point Likert scale ranging from ‘strongly agree=1’ to ‘strongly disagree=5’.

The final categorical status of each participant was determined by collapsing the ‘definite’ and ‘probable’ categories of the experts’ responses to create a dichotomous variable and then using a majority (2 out of 3) or better (3 out of 3) agreement among the three expert judges to determine the final status (Kim et al., 2001). Owing to unavoidable circumstances, two of the three original raters were not able to complete all of the interviews. However, we had a back-up expert judge (a psychiatrist trained in both internal medicine and psychiatry, who primarily works with people with schizophrenia) whose scores were used whenever there was a missing judgement among the first three judges. The experts rendered their categorical judgements independently of one another.

Categorical capacity judgements were rendered for 101 participants: 90 with severe mental illness and 11 controls. The videotape for one participant with mental illness was not used because the sound quality was poor. Moreover, because of lower variance with higher performance (ceiling effect) in the comparison group, we only used 11 of the 40 tapes, including the two lowest scoring control participants. Expert judge 1 reviewed all 101 interviews, judge 2 reviewed 79 interviews and judge 3 reviewed 91 interviews. There were no participants who had a missing judgement from more than one judge. The back-up expert judge rendered judgements for 72 interviews and, of those, 32 judgements in which there was a missing judgement from either judge 2 or judge 3 were used in the final determination of capacity status (the back-up judge rated more than the 32 participants with missing ratings to assess reliability among all four judges).

The rationale and methodology for the expert judgement criterion method have been described elsewhere (Kim, 2006), including its advantages over an a priori cut-off criterion (Wirshing et al., 1998; Moser et al., 2002) and a psychometric criterion (Marson et al., 1995; Grisso et al., 1997; Schmand et al., 1999; Kovnick et al., 2003). Given that most societies look to clinicians’ judgement about such decisions, expert judgement offers an arguably more valid standard against which to measure participant performance. Methodologically, expert judgement provides an independent assessment criterion, since the experts are not affiliated with the schizophrenia research studies.

**Statistical analyses**

Group comparisons of demographic, symptom severity and MacCAT–CR summary data were conducted using parametric or non-parametric tests. Pairwise and group
kappa coefficients were calculated to assess categorical agreement among expert judges. Receiver operating characteristic (ROC) analysis was conducted to assess the test characteristics of each of the three subscales (understanding, appreciation, and reasoning) of MacCAT–CR against the final categorical judgements made by the expert judges. To demonstrate how the sensitivity and specificity data generated from the ROC analysis can be applied to potential research scenarios, we calculated the positive and negative predictive values (PPVs and NPVs) for a range of hypothetical prior probabilities for three cut-off points on the understanding sub-scale.

Data were analysed using SPSS version 12.0 and Stata version 8.0 (both for Windows).

RESULTS

The group with severe mental illness and controls showed no significant differences in age, gender and race distribution, or educational level (Table 1). None of the controls had a psychotic disorder; 6 had a mood disorder, 1 a substance dependence disorder and 1 an anxiety disorder. Seventy-five of the group with severe mental illness had schizophrenia, 14 had schizoaffective disorder and 2 had affective disorders with psychosis. Among the 55 participants from the CATIE study, 49 had schizophrenia and 6 had schizoaffective disorder.

Performance on MacCAT–CR

Those with severe mental illness performed significantly worse than the comparison group on the MacCAT–CR sub-scales (except for choice). Within this group, the 55 participants from the CATIE study performed better than the other participants on all sub-scales of the MacCAT–CR: understanding, mean score (s.d.) 21.3 (3.7) vs. 19.1 (5.6), t = 2.1, d.f. = 54.7, P = 0.04; appreciation, 4.1 (1.5) vs. 3.4 (1.8), t = 2.0, d.f. = 66.0, P = 0.05; reasoning, 5.4 (1.5) vs. 4.8 (2.3), t = 1.4, d.f. = 54.2, P = 0.17. This is consistent with our goal of avoiding spectrum bias by expanding the range of scores in the group with severe mental illness.

Expert judgements

Of the 101 CATIE participants reviewed, 25 (including 7 of the 55 CATIE participants) were deemed probably or definitely incapable of consent. The pairwise kappa coefficients among the four judges ranged from 0.56 to 0.90; the group kappa coefficient for the four expert judgements was 0.69 (Z = 14.1, P < 0.001). When asked whether or not the videos provided a sufficient basis for them to make their capacity determinations, the mean rating ranged between strongly agree = 1 and agree = 2 for three of the experts, with mean (s.d.) ratings of 1.4 (0.9), 1.4 (0.8) and 1.9 (0.9), and between agree = 2 and neutral = 3 for the remaining expert judge, whose mean rating was 2.5 (1.1).

Predictive values of MacCAT–CR scores

Table 2 summarises the sensitivity and specificity using various cut-off points on the three sub-scales of the MacCAT–CR. The area under the ROC curve was higher for the understanding sub-scale at 0.94 (95% CI 0.88–0.99) than for the appreciation sub-scale (0.85, 0.76–0.94) and the reasoning sub-scale (0.80, 0.70–0.90), indicating that MacCAT–CR scores, especially for understanding, were significant predictors of categorical capacity status. However, none of the sub-scales had a single cut-off score with a very high sensitivity and specificity.

Sensitivity and specificity are features of tests, not populations, and cannot guide decisions without information about prevalence. For the purpose of determining the acceptable capacity scores that might be recommended (for instance to a research ethics committee reviewing a research protocol to be used to screen people with impaired capacity), the results of the ROC analysis were used to generate positive predictive values (PPV, the probability that a person found to perform at or below a MacCAT–CR sub-scale cut-off score will in fact be incapable) and negative predictive values (NPV, the probability that a person performing above the cut-point will in fact be capable), as shown in Table 3.

A high PPV implies a low false-positive rate (i.e. low likelihood of mistakenly excluding a capable person); a high NPV implies a low false-negative rate (i.e. low likelihood of mistakenly enrolling an incapable person). In determining what degree of decisional capacity to require of research participants, it would be undesirable to use a high cut-off score when prevalence is low (e.g. understanding score of 21 at 10% prevalence of incapacity) because 76% of persons excluded as too impaired will in fact be capable (given the PPV of 24%). Such a practice would not only be inefficient but also would unfairly exclude willing and capable persons from participating in research. It would also be

Table 1 Participants’ characteristics and performance on the MacArthur Competence Assessment Tool–Clinical Research

<table>
<thead>
<tr>
<th></th>
<th>Participants with severe mental illness (n=91)</th>
<th>Controls (n=40)</th>
<th>Test statistic</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female gender, n (%)</td>
<td>Female gender, n (%)</td>
<td></td>
<td>χ²</td>
<td>t</td>
</tr>
<tr>
<td>Black or minority, n (%)</td>
<td>Black or minority, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, years: mean (s.d.)</td>
<td>Age, years: mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education, years: mean (s.d.)</td>
<td>Education, years: mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PANSS score: mean (s.d.)</td>
<td>PANSS score: mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MacCAT–CR score: mean (s.d.)</td>
<td>MacCAT–CR score: mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding</td>
<td>Understanding</td>
<td></td>
<td>−7.4</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Appreciation</td>
<td>Appreciation</td>
<td></td>
<td>−5.3</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Reasoning</td>
<td>Reasoning</td>
<td></td>
<td>−6.2</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Choice</td>
<td>Choice</td>
<td></td>
<td>−1.8</td>
<td>0.08</td>
</tr>
</tbody>
</table>

2. The CATIE participants’ scores were lower (indicating a lower degree of psychopathology) than other participants with severe mental illness: positive, 14.7 x 19.8 (P < 0.001); negative, 17.4 x 19.6 (P = 0.07); general, 33.2 x 43.1 (P < 0.001).
### Table 2: Sensitivity and specificity of cut-off scores on sub-scales of the MacArthur Competence Assessment Tool—Clinical Research

<table>
<thead>
<tr>
<th>MacCAT–CR sub-scale</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding (scale range 0–26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>6</td>
<td>0.040</td>
<td>1.000</td>
</tr>
<tr>
<td>9</td>
<td>0.080</td>
<td>1.000</td>
</tr>
<tr>
<td>10</td>
<td>0.160</td>
<td>1.000</td>
</tr>
<tr>
<td>11</td>
<td>0.200</td>
<td>1.000</td>
</tr>
<tr>
<td>12</td>
<td>0.240</td>
<td>1.000</td>
</tr>
<tr>
<td>13</td>
<td>0.360</td>
<td>1.000</td>
</tr>
<tr>
<td>14</td>
<td>0.400</td>
<td>1.000</td>
</tr>
<tr>
<td>15</td>
<td>0.560</td>
<td>1.000</td>
</tr>
<tr>
<td>16</td>
<td>0.680</td>
<td>1.000</td>
</tr>
<tr>
<td>17</td>
<td>0.760</td>
<td>0.961</td>
</tr>
<tr>
<td>18</td>
<td>0.800</td>
<td>0.921</td>
</tr>
<tr>
<td>19</td>
<td>0.840</td>
<td>0.803</td>
</tr>
<tr>
<td>20</td>
<td>0.840</td>
<td>0.800</td>
</tr>
<tr>
<td>21</td>
<td>0.880</td>
<td>0.684</td>
</tr>
<tr>
<td>22</td>
<td>0.960</td>
<td>0.618</td>
</tr>
<tr>
<td>23</td>
<td>1.000</td>
<td>0.474</td>
</tr>
<tr>
<td>24</td>
<td>1.000</td>
<td>0.368</td>
</tr>
<tr>
<td>25</td>
<td>1.000</td>
<td>0.171</td>
</tr>
<tr>
<td>26</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Appreciation (scale range 0–6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.080</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>0.360</td>
<td>0.987</td>
</tr>
<tr>
<td>2</td>
<td>0.480</td>
<td>0.921</td>
</tr>
<tr>
<td>3</td>
<td>0.840</td>
<td>0.750</td>
</tr>
<tr>
<td>4</td>
<td>0.920</td>
<td>0.553</td>
</tr>
<tr>
<td>5</td>
<td>0.960</td>
<td>0.276</td>
</tr>
<tr>
<td>6</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Reasoning (scale range 0–8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.120</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>0.160</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>0.200</td>
<td>0.974</td>
</tr>
<tr>
<td>3</td>
<td>0.360</td>
<td>0.947</td>
</tr>
<tr>
<td>4</td>
<td>0.600</td>
<td>0.842</td>
</tr>
<tr>
<td>5</td>
<td>0.840</td>
<td>0.592</td>
</tr>
<tr>
<td>6</td>
<td>0.960</td>
<td>0.368</td>
</tr>
<tr>
<td>7</td>
<td>0.960</td>
<td>0.171</td>
</tr>
<tr>
<td>8</td>
<td>1.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

1. Categorical capacity status determined by expert re-views of 80 participants (90 with severe mental illness and 11 controls).
2. Of those deemed incapable of providing consent, the proportion who scored at or below the cut-off score.
3. Of those deemed capable of providing consent, the proportion who scored above the cut-off score.

### Table 3: Positive and negative predictive values for three potential cut-off scores on the MacArthur Competence Assessment Tool—Clinical Research understanding sub-scale, for a range of prevalence values

<table>
<thead>
<tr>
<th>Prevalence of incapacity, %</th>
<th>Cut-off score 15</th>
<th>Cut-off score 18</th>
<th>Cut-off score 21</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PPV, %</td>
<td>NPV, %</td>
<td>PPV, %</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>95</td>
<td>53</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>90</td>
<td>72</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>84</td>
<td>81</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>77</td>
<td>87</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>69</td>
<td>91</td>
</tr>
</tbody>
</table>

PPV, positive predictive value; NPV, negative predictive value.
1. Sensitivity 0.56, specificity 1.0.
2. Sensitivity 0.80, specificity 0.92.
3. Sensitivity 0.88, specificity 0.68.

### DISCUSSION

Despite increasing research on the decision-making abilities of people with neuro-psychiatric disorders, there are few data on how to translate information about impairment into categorical determinations. In the real world, it is necessary to determine the categorical capacity status of the potential participant, i.e. whether they are capable of providing informed consent. This information is needed for excluding those who are incapable, for identifying those in need of surrogate decision-makers, or for identifying those who may require remedial education. Thus, an important goal of capacity research in schizophrenia is to inform policies and practices that help guide the determination of categorical capacity status of potential participants (Kim, 2006).

In the research context, informed consent disclosures are relatively consistent across participants, since the relevant information, including the risk–benefit ratio, is determined by the characteristics of a research protocol which is applicable to all potential participants. This is in contrast to the treatment context in which the procedures, risks, benefits and hence disclosures might be unique to each individual's treatment situation, and for whom the assessment of decision-making capacity requires individualised patient information (Cairns et al, 2005a). Further, whereas in the treatment context the welfare of the patient is the physician's paramount concern, in the research context, the investigator's priority is the advancement of science, thus increasing the need for a more transparent and objective process for determination of capacity. Therefore, the research context provides an opportunity as well as an imperative to create a standardised capacity assessment by using an assessment instrument that can be benchmarked against ethically appropriate, methodologically rigorous independent validation provided by experienced clinicians.

### Objective determination of capacity for research

Our study establishes the feasibility of an objective assessment of capacity for the research context. By validating the MacCAT–CR sub-scales against an expert judgement standard, we can go beyond mere descriptions of participants’ performance on a scale. For example, given that the prior probability of incapacity among those screened for the CATIE–Schizophrenia study was probably quite low (Stroup et al, 2005), we can surmise that even a low cut-off score on the understanding sub-scale such as 15 (which was in fact used by the CATIE study) would rarely include people lacking capacity (e.g. at an estimate of 10% prevalence, there would only be a 5% false-negative rate), with virtually no chance of mistakenly identifying those with capacity as incapable.

Since most research studies could probably be categorised into a handful of categories in terms of their risk–benefit ratio (Maryland Attorney General’s Research Working Group, 1998; National Bioethics...
Advisory Commission, 1998; New York Department of Health Advisory Work Group on Human Subject Research Involving the Protected Classes, 1999), a limited number of targeted validation studies would likely provide a sufficient evidence base for ethically appropriate yet efficient practice for a variety of research studies involving participants with impairment in decision-making. In effect, a series of tables such as Table 3 could provide a systematic and objective guide to research ethics committees and investigators.

How important is the fact that the sub-scales of the MacCAT–CR do not seem to have a single cut-off point that has both high sensitivity and specificity? First, it might simply be unrealistic to expect extreme precision and predictability from a standardised instrument when it is applied to making complex, value-laden judgements about a person’s decision-making capacity. Second, this limitation might not be a problem as long as the purpose of assessing capacity is clear; for instance, one might focus more heavily on the PPV or the NPV, depending on the situation. So for an early-phase study of an invasive intervention likely to yield no benefit to the participants but which may pose some risk, the thresholds could be set high enough to eliminate any one who lacks capacity (false-negatives). Alternatively, if such a method eliminates too many potential participants as false-positives, a two-step approach could be used: a lower MacCAT–CR threshold to decrease false-positives and then individual in-depth capacity assessments to ensure that only competent persons are enrolled. As this last example illustrates, we believe that our method should be used flexibly to meet the ethical requirements of the situation, rather than rigidly adhering to a formula.

Strengths
To our knowledge, this is the most thorough validation against expert judgement of capacity thresholds on a widely used capacity assessment tool for clinical research, and the first such validation study involving people with schizophrenia. This study has several strengths. The majority of participants with severe mental illness were in an actual clinical trial. This group exhibited a wide spectrum of impairment, allowing us to conduct a meaningful ROC analysis. The expert judges achieved relatively high levels of non-chance agreement, and they felt that in general they had sufficient information to make the capacity determinations. The expert judges based their judgements on video recordings of interviews, which provided more information than written transcripts.

Limitations
There are however some limitations to the study and caveats. First, our sample consisted of both CATIE participants and people with severe mental illness who were not in the CATIE study. The latter were included to ensure a sufficient spectrum of performance for the ROC analysis. Thus, no generalisations regarding the relative performance of the subgroups should be drawn from our data. The CATIE participants might have performed better on the MacCAT–CR because they had a less severe illness but also because, being involved in the study, the study protocol had been previously explained to them in more depth.

Second, the experts’ judgements were based on their viewing the taped MacCAT–CR interviews rather than performing their own independent assessments (which was not feasible in this multisite study involving multiple expert judges). Thus, our method is susceptible to incorporation bias that can falsely increase the accuracy of the test (Zhou et al, 2002). However, this limitation must be weighed against the following countervailing considerations. Currently, there is a lack of standardised procedures for capacity determinations. The MacCAT–CR covers the essential elements of a capacity assessment (Appelbaum & Grisso, 2001) and its standardised nature mitigates the variability of capacity assessments. In the absence of agreed procedures for capacity assessments, a criterion standard based on various experts’ evaluations (even if it were feasible in a multisite study such as this) would involve a variety of methods, creating uncertainties regarding the nature of the standards used. Thus, although we cannot rule out the possibility of incorporation bias, we believe our results represent a reasonable balance between feasibility and validity.

Third, before the results of our study are generalised to other contexts, one must take into account the potential adverse effects of focusing on ‘cut-off scores’ of capacity assessment instruments (Grisso & Appelbaum, 1996), especially the danger that the cut-off scores will be seen as inherent features of the assessment instrument (i.e. anyone scoring above a certain level has adequate capacity for any decision), rather than needing context-by-context validation and context-sensitive application. To avoid such misuse, any generalisation of our validation method must take into account two points.

First, the prevalence of incapacity in schizophrenia studies other than the CATIE study might be different for a variety of reasons. For instance, in studies that target people with refractory illness or those who are long-term in-patients (Kovnick et al, 2003), the prevalence rates will be higher and the estimation of PPV and NPV will need to take that into account. Second, any attempt to generalise our validation method to other schizophrenia studies must take into account the risk-sensitive nature of capacity thresholds (Brock, 1991; Grisso & Appelbaum, 1998; National Bioethics Advisory Commission, 1998). In our study, the expert raters made judgements regarding the level of capacity that was adequate for a relatively low-risk clinical trial. However, the risk–benefit ratio might be different for studies involving placebos, symptom provocation, or phase I tests of invasive interventions. The ROC curves for such studies might look quite different from those in this study.

Finally, the fact that 7 of 55 CATIE study participants were deemed to lack capacity by our experts needs to be interpreted with caution. Our CATIE sample was not intended to be representative of the overall CATIE study. The ratings of the expert judges were not available to the CATIE investigators at the time that they made their judgements regarding admission to the CATIE study. Further, a number of unique safeguards (Stroup et al, 2005), including independent participant advocates (Stroup & Appelbaum, 2003), were built into the CATIE project. Finally, in the absence of a true ‘gold standard’ for determining categorical status, we are proposing the expert judgement-based method as a provisional criterion standard that needs to be further studied and improved (Kim, 2006).

Future directions
The results of our study provide an evidence-based decision framework for how to use instruments for measuring
decisional abilities to guide valid categori- cal judgements about a potential partici- pant’s capacity to give informed consent. We believe that as long as its limitations and caveats are kept in mind, future re- search employing the framework provided in our study could have important practical implications. By performing validation studies for a few categories of risk–benefit situations, it might even be possible to interpolate reasonable guidelines for most schizophrenia research studies. Such an approach would make the crucial task of determining a potential participant’s capacity status much more transparent, objec- tive and evidence-based than it is today.

ACKNOWLEDGEMENTS

We thank Linda Ryan, MD, Lior Givon, MD and Telva Olivarres, MD for their expert ratings, Sonia Davis, PhD for assistance with database management and Jayendra Patel, MD for assistance with recruitment. The study was supported by the National Institute of Mental Health USA (grants K23 MH64172 and N01 MH90001).

REFERENCES


(First received 13 November 2006, final revision 13 March 2007, accepted 23 March 2007)
Young people who self-harm

ROBERT YOUNG, MICHAEL VAN BEINUM, HELEN SWEETING and PATRICK WEST

Background  Self-harm among young people in the UK is possibly increasing but little is known about the reasons young people give for cessation and their link with gender or employment status.

Aims  To investigate self-harm in young people, prevalence, methods used, motivations for starting and ceasing, service use, and how these are related to gender, parental social class and current labour market position.


Results  Both past and current rates of self-harm were highest among those outside the labour market. This group was most likely to want to kill themselves and did not cite specialist mental health services as helpful in ceasing self-harm. Those in full-time education more often self-harmed for a brief time, mainly to reduce anxiety.

Conclusions  Current labour market position was a stronger predictor than parental social class or gender for self-harm, and was linked to level of severity, motivation for starting and ceasing, and service utilisation.

Declaration of interest  None.

Funding detailed in Acknowledgements.

Self-harm is a relatively common and possibly increasing (Gunnell et al, 2000) problem among young people in the UK, affecting 7–14% at some point in their life (Hawton et al, 2002; Hawton & James, 2003; Skegg, 2005). The majority, however, do not present to statutory services (Hawton et al, 2002; Nada-Raja et al, 2003). Considerable research has explored self-harm behaviour in young people, including examination of socio-economic status, gender and individual factors such as sexual orientation and identity (Platt & Hawton, 2002; De Leo & Heller, 2004; Skegg et al, 2003; Reikopf & Buka, 2006; Young et al, 2006). There is, however, only a single peer-reviewed study exploring the reasons for ceasing self-harm (Sinclair & Green, 2005), but potential risk factors such as gender or employment status were not investigated.

This population-based survey of 18- to 20-year-olds investigated three well-established predictors of self-harm – gender, parental social class and current labour market position – and examined how they relate to reasons for both self-harm behaviour and its cessation, and the use of social supports.

METHOD

Participants  Participants were 1258 (49% of baseline sample) 18- to 20-year-olds from the longitudinal West of Scotland 16+ Study of health and lifestyles (Sweeting et al, 2005). They were originally recruited during their final year (1994) of primary school (age 11, n=2586, 93% of issued sample) and surveyed under exam-type conditions using questionnaires at ages 11, 13 and 15 years, and by personal interview at 18–20 years (2002–2004). Losses to follow-up were typical (e.g. more likely to originate from households of lower social class, have low educational involvement and belong to reconstituted or lone-parent households; Sweeting et al, 2001, 2005). As weights to adjust for attrition bias did not alter the results, we report unweighted data (weighted results are available from the authors). Participants were interviewed individually at 18–20 years of age by registered nurses using computer-assisted interviews, either at a central location (Glasgow University or participants’ old secondary schools) or at home, depending upon availability and preference. The study received ethical approval from the Glasgow University Ethics Committee.

Measures  All measures, apart from parental social class – which was obtained at age 11 – were determined at age 18–20. Questions about self-harm were asked in a single section of the interview in the sequence given below.

Lifetime self-harm and method(s)  All participants were asked ‘Have you ever tried to hurt yourself or harm yourself deliberately?’ and, if yes, what method(s) they had used from the following list: cutting (on the arm or wrists); cutting elsewhere on the body; scratching or scoring; taking dangerous tablets or pills; hitting or punching self; slamming hands in door; burning (with cigarettes, lighter, etc); other way (please specify).

Reason for self-harming and age at onset  Participants who had self-harmed were asked ‘What are / were the reasons for doing this?’ from the following list: to upset others; relieve anxiety; relieve anger; forget about something; make someone else take notice; punish myself; kill myself; not sure why; other reason (please specify). In addition, they were asked at what age they first started to harm themselves.

Current self-harm and awareness by health services and informal networks  Those who reported self-harm were asked whether this was in the past only, currently (in the past year) only, or past and current. Anyone who had self-harmed within the past year was asked ‘Who currently knows about this?’ from a list including: psychiatrist or other mental health professional; doctor/general practitioner (GP); parents (either); spouse or partner; friend(s); brother(s)/sister(s) or other close family
member; work or college; most people that know you fairly well; none of the above. Responses were subsequently collapsed into formal services (GP, psychiatrist and health professional) and informal networks (parents, family, spouse or friends).

**Reasons for stopping**

Those who admitted to self-harm but said that they were not currently doing so were asked the open response question: 'Why did you stop?' We categorised the 65 valid responses into four broad categories that show considerable similarity to the narratives developed by Sinclair & Green (2005). The categories were: one off or temporary phase (e.g. 'only happened once, no further thoughts of self-harm'); coped or felt better or found purpose (e.g. 'felt I was coping better with things and that [there] were better ways to cope'); got professional help or help from family or friends (e.g. 'went to see psychiatrist'); realised the harm to self or family or the 'stupidity' of self-harm (e.g. 'realised what life was worth and how much it hurt the family').

**Service use**

To assess service use, participants were asked to select from a list of services those they had used since age 11. We report use of Scottish services most relevant to self-harm, namely psychiatric, accident and emergency, Children’s Panel (part of the Scottish Youth Justice System) and social work.

**Parental social class**

Parental social class was based on the occupation of the head of the household (father figure’s current or previous occupation), mainly provided by parents during the first wave of the study (when the participants were aged 11 in 1994). Where missing, these data were supplemented by information provided by the children themselves, which we have found to be reliable (West et al., 2001). Occupations were categorised by reference to the 1990 Standard Occupational Classification (Office of Population Census and Surveys, 1991) into non-manual (occupational classes I–IIIInm) and manual (classes IIIm–V). There were 63 instances in which social class was unclassifiable owing to either missing or poor information. With the exception of basic statistics, unclassifiable data for social class were treated as missing.

**Current labour market position**

Participants were asked a set of questions concerning education, training and employment to determine their main labour market position. This was classified into three broad groups: full-time education (higher or further education); training or work (either full- or part-time, or on a training course or scheme); and a non-labour market group, comprising unemployed (n=86), at home or with care responsibilities (n=37) and those sick or ill (n=12).

**Statistical analysis**

Chi-squared or Fisher's exact tests were used, as appropriate, for categorical data, and two-tailed t-tests to assess differences for age at onset. Logistic regression was used to test for potential confounding between parental social class and current economic position. Re-analysis of data omitting the 12 participants classified as sick or ill did not alter the results substantially. Owing to the relatively low frequencies and their more qualitative nature, we report only the raw numbers for some of the more exploratory analyses of young people’s explanations for ceasing self-harm.

**RESULTS**

**Self-harm, gender and labour market position**

Table 1 shows the rates of current, past and lifetime self-harm according to gender, parental social class and current labour market position. Overall, we found a 7.1% lifetime prevalence of self-harm, with the majority self-harming in the past only. Only 1.6% were currently self-harming. Despite no statistically significant gender difference, there was a suggestion that young women were more likely to self-harm during their lifetime (8.4 vs. 5.8%). Parental social class did not predict self-harm, but current labour market position was strongly related. There was a threefold increase in lifetime self-harm, and a six- to sevenfold increase for current self-harm, for the non-labour market group compared with those in work or full-time education. In a logistic regression, including all three socio-demographic variables (female gender OR=1.52, 95% CI 0.97–2.40; parental social class – manual OR=0.69, 95% CI 0.42–1.11; and current labour market position – full-time education v. training or work OR=1.47, 95% CI 0.86–2.52, full-time education v. non-labour market OR=3.93 95% CI 2.12–7.29), only current labour market position remained a significant predictor of lifetime self-harm. Young women were more likely to report starting to self-harm at an earlier age (females: mean 15.0, s.d.=2.1 years; males: mean 16.4, s.d.=2.3; two-tailed t-test

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Current and past self-harm according to gender, parental social class and current labour market position</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-harm</strong></td>
<td><strong>Gender</strong></td>
</tr>
<tr>
<td>Male (n=640)</td>
<td>Female (n=618)</td>
</tr>
<tr>
<td>Never, n (%)</td>
<td>603 (94.2)</td>
</tr>
<tr>
<td>In past only, n (%)</td>
<td>28 (4.4)</td>
</tr>
<tr>
<td>Current (in past year), n (%)</td>
<td>8 (1.3)</td>
</tr>
<tr>
<td>Ever, n (%)</td>
<td>37 (5.8)</td>
</tr>
</tbody>
</table>

1. One participant refused to answer about current self-harm, and because of this and rounding errors column totals may not be 100%.
2. Current labour market position difference in self-harm (never: p<0.001).


Table 2  Method of self-harm according to gender, parental social class and current labour market position

<table>
<thead>
<tr>
<th>Method of self-harm</th>
<th>Gender</th>
<th>Parental social class</th>
<th>Current labour market position</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (n=640)</td>
<td>Female (n=618)</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>Non-manual (n=611)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Manual (n=584)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Any cut, self-scoring or scratching</td>
<td>17 (2.7)</td>
<td>38 (5.7)</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>Cutting (on the arm or wrists)</td>
<td>14 (2.2)</td>
<td>32 (5.2)</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Cutting (elsewhere on the body)</td>
<td>5 (0.8)</td>
<td>10 (1.6)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Scratching or scoring</td>
<td>5 (0.6)</td>
<td>6 (1.0)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Taking dangerous pills</td>
<td>8 (1.3)</td>
<td>27 (4.4)</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Hitting or punching self</td>
<td>10 (1.6)</td>
<td>5 (0.8)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Burning (with cigarette, lighter, etc.)</td>
<td>6 (0.9)</td>
<td>3 (0.5)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Slapping hand, etc. in door</td>
<td>4 (0.6)</td>
<td>4 (0.6)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Others²</td>
<td>5 (0.8)</td>
<td>6 (1.0)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17 (2.7)</td>
<td>38 (5.7)</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14 (2.2)</td>
<td>32 (5.2)</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 (0.8)</td>
<td>10 (1.6)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 (0.6)</td>
<td>6 (1.0)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 (1.3)</td>
<td>27 (4.4)</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 (1.6)</td>
<td>5 (0.8)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 (0.9)</td>
<td>3 (0.5)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 (0.6)</td>
<td>4 (0.6)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 (0.8)</td>
<td>6 (1.0)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26 (3.9)</td>
<td>12 (2.6)</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23 (3.5)</td>
<td>10 (2.2)</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 (1.2)</td>
<td>4 (0.9)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 (0.9)</td>
<td>1 (0.2)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 (0.8)</td>
<td>1 (0.2)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 (0.5)</td>
<td>3 (0.6)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 (0.2)</td>
<td>9 (1.5)</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 (0.8)</td>
<td>1 (0.2)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 (0.5)</td>
<td>3 (0.6)</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18 (3.1)</td>
<td>18 (3.1)</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 (1.5)</td>
<td>2 (1.5)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 (2.2)</td>
<td>3 (2.2)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9 (0.7)</td>
<td>9 (0.7)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 (0.6)</td>
<td>8 (0.6)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11 (0.9)</td>
<td>11 (0.9)</td>
<td>NS</td>
<td></td>
</tr>
</tbody>
</table>

1. Sixty-three omitted because of undistinguishable social class.
2. Includes self-poisoning and hanging, self-suffocation and hanging, use of inhalant and taking pills, alcohol to provoke stomach ulcer, punching a wall, no answer.

Methods of self-harm

Table 2 shows rates of self-harm by different methods within the sample as a whole. Cutting, scoring or scratching were the most common, followed by taking dangerous tablets; other (typically overtly violent) methods such as burning or punching self were relatively rare. A clear gender pattern emerged, with young women more likely to cut themselves or take dangerous tablets. The only difference according to parental social class was that those from manual backgrounds were more likely to self-harm than those in non-manual backgrounds (35.9%) being much more likely than those from non-manual backgrounds (6.7%) to self-harm to kill themselves. Those from non-manual backgrounds were also more likely to be unclear about why they had harmed themselves. Those in full-time education were more likely to self-harm to reduce anxiety than those in training or work or the non-labour market group (although not significantly so), and those in the non-labour group were more likely to self-harm in order to kill themselves.

Reasons for self-harm

Table 3 shows the reasons cited by those who had self-harmed. Relief of anxiety was most commonly reported, followed by wanting to forget about something, relief of anxiety and desire to kill themselves. This confirms that the main motive behind young people’s self-harm was to relieve negative emotions, with only a small minority saying they self-harmed in order to elicit attention and help from others.

People aware of young person’s self-harm

The small numbers of those who were currently self-harming (n=20) made it difficult to establish statistical patterns. According to the young people’s reports, at least one person or agency was aware of their behaviour in most cases (17 out of 20). All mentioned somebody from their informal network, most usually a parent (11 out of 20). Formal services were less likely to be aware of the behaviour (9 out of 20), with GPs more likely (9 out of 20) than specialist mental health professionals (4 out of 20) to know; no young person had told their college or work.

There were no gender differences with respect to who was aware of the self-harm but, despite the small numbers, we found significant differences in parental social class and current labour market position. Those from households with a manual worker as the head or outside the labour market were more likely to confide in a friend. Those outside the labour market were also more likely to tell a psychiatrist or mental health professional about their self-harm; indeed, none of those in full-time education or in work or training had confided in a psychiatrist or mental health professional (further details available from the authors).

Explanations for stopping self-harm

Table 4 shows the main explanations given by participants as to why they had stopped

$t=2.9, d.f.=84, P=0.004$, but there was no significant difference in age at onset according to parental social class or current labour market position.
professionals, close friends or an unspecified purpose in life. The least common explanation (or 'stupidity') was strongly associated with reason for ceasing self-harm, although the small numbers in each category suggest caution in interpretation. Those in full-time education were more likely to attribute their self-harm to a temporary phase, particularly compared with those outside the labour market. Those in training or work were more likely to attribute stopping their self-harm to 'coping better'. Half the young people outside the labour market attributed either upset to family, friends or dependants, or realising the futility/stupidity of harm to self as reasons.

### Service use

Overall, no gender differences were found in service use, and the only difference according to social class was that those from a manual background were more likely to present to an accident and emergency department. Young people currently outside the labour market appeared to be the highest users of statutory services, with elevated use of psychiatric, accident and emergency, Children’s Panel and social work services.
DISCUSSION

Previous studies have indicated that gender, social class and labour market position are all important predictors of self-harm (Platt & Hawton, 2002; Skegg, 2005). This study reports on variation in self-harm in relation to these three socio-demographic factors and investigates how they may be linked to the explanations given for engaging in and stopping self-harm, as well as considering implications for service provision. The prevalence estimates in this study are similar to those from other population-based studies (Hawton et al., 2002; Hawton & James, 2005; Skegg, 2005), strengthening the generalisability of our findings. There has only been one previous peer-reviewed study of reasons for cessation in a population who had previously self-harmed (Sinclair & Green, 2005). Our study, as well as confirming the four key narratives described in that clinic study, was able to link reasons for stopping self-harm with gender and current labour market position.

We found that the main motive behind most young people’s self-harm was to relieve negative emotions. This is consistent with the few population-based studies, which have suggested that young people who self-harm may have limited coping strategies to deal with emotional difficulties or may be exposed to elevated stress levels. For example, the Child and Adolescent Self-Harm in Europe Study found that the most common immediate reason for self-harm was ‘to find relief from a terrible state of mind’, or ‘wanting to die’, but other, less common, reasons included ‘to punish self’ and to bring their distress to others’ attention (Rodham et al., 2004). Similar reasons are also given by clinic attendees (Nock & Prinstein, 2004).

Previous studies have identified gender as an important predictor of self-harm, with a higher prevalence among young women (Hawton et al., 2002; Hawton & James, 2005; Skegg, 2005). Although our study found an excess of females for lifetime self-harm, current rates were similar for both genders. Young women were more likely to state that they would self-harm to reduce anxiety but, counter to traditional gender differences, reducing anger via self-harm was unrelated to gender. In addition, young women were more likely to self-harm by cutting or taking tablets, whereas young men were more likely to use violent methods, as in other studies (Lewinsohn et al., 1996). We found no gender difference in relation to the number of young people reporting that they self-harmed to kill themselves. This contrasts with reported gender differences in suicide rates (Skegg, 2005) and may indicate that gender differences in completed suicides could be partially attributed to gender differences in the lethality of their chosen methods of self-harm. None of the young men said that either professional or more informal help was a primary factor in stopping self-harm, but 8 out of 40 young women said this was the main reason for cessation. One possible explanation is that current professional therapeutic interventions are more tailored towards women, who may find it easier to discuss emotional difficulties than men (van Beinum, 2003; Biddle et al., 2004).

Previous studies have shown that rates of suicide, attempted suicide and self-harm are related to socio-economic factors, although the relationship is by no means straightforward (Platt & Hawton, 2002; Rehkopf & Buka, 2006). In our study, contrary to previous research, the overall prevalence of self-harm was not strongly related to parental social class, although one specific reason (‘killing myself’) was more often cited by young people from manual social backgrounds. The lack of association with social class is unexpected but is consistent with evidence of equalisation in health among young people in contemporary society (West & Sweeting, 2004). When we compared the relative effect of parental social class with current labour market position, we consistently identified current labour market position as the more important factor for self-harm. Young people most at risk were those who were currently unemployed, sick or outside the labour market. This closely mirrors the results of a previous study of the impact of youth unemployment on suicidal behaviour among 18-year-olds in the same geographical area (West, & Sweeting, 1996).

Other results confirm the greater severity of self-harm among those outside the labour market, with nearly half (10 out of 23) explaining that their reason for self-harm was to kill themselves, and many reporting high service use, particularly of mental health services. However, service use may not have been related solely to self-harm, as this group are likely to have other psychological or behavioural problems that increase their use of statutory services. Few outside the labour market attributed their self-harm to a transitory phase, indicating a chronic problem, and none of the 12 who had ceased to self-harm said that specialist health services were useful in supporting them to stop.

For young people in education self-harm was more likely to be a transitory reaction to specific stress, such as examinations or academic pressures, which have previously been related to psychological distress in this cohort (West & Sweeting, 2003). For many in this group, self-harm might have been an adaptive coping mechanism to deal with temporary anxiety states and not something for which they felt they needed external help. This is compatible with research suggesting that young people who self-harm are less likely to use other coping strategies in times of stress (Evans et al., 2005). This is supported by a recent study which found that young people with anxiety disorders were least likely to use statutory services (Ford et al., 2006). Furthermore, those in our study who were currently self-harming, and who were in education, work or training, tended to be more secretive, nearly always concealing their behaviour from professionals, parents or friends, and not disclosing problems at work or college.

Study limitations

This study is restricted to 18- to 20-year-olds and therefore conclusions about self-harm in other age groups are not possible. Glasgow has both a relatively high level of deprivation and a high concentration of colleges and universities, and this may have boosted the power of this study to detect differences between students, those employed and those outside the labour market. Attrition may also be important, since it tends to disproportionally affect those from disadvantaged backgrounds. However, applying weights to adjust for this had negligible effects on the results, suggesting it was not a factor. The study relies on young people’s reports of self-harm, but this itself may be socially patterned. Owing to the relative simplicity of this analysis there is always the possibility of omitting relevant variables. For instance, the small numbers did not allow investigation of the impact of affective and psychotic disorders on self-harm. Similarly, those outside the labour market were a heterogeneous group but their small numbers did not allow for a more detailed investigation of possible subgroup differences. However, despite these limitations, this study represents one of
the largest population-based studies of self-harm in this age group.

Implications

The transient nature of self-harm behaviour found in young people in employment or education suggests a better clinical outcome for this group, despite their reluctance to access help. The results reported here suggest that the most acceptable supports for these two groups would be approaches that emphasise developing personal coping skills. In contrast, those young people who are unemployed, sick or not in full-time education are of greater concern. They are more likely to be engaging in chronic self-harm and to be actively trying to kill themselves. An important finding was that, in our study, 45% of young people who had self-harmed were known to their GP, compared with 4% in Australia (De Leo & Heller, 2004) and 13% in New Zealand (Nada-Raja et al, 2003). Therefore, Scottish GPs might be a means of targeting intervention. However, although a number of respondents said that they had accessed specialist services, they often had not found them particularly helpful. More effective interventions for this group of vulnerable young people are urgently required and, in particular, training and additional support for GPs. Targeting upstream causes (social disadvantage and chaotic personal circumstances) are likely to prove more effective than biomedical interventions alone (Platt et al, 2005).

ACKNOWLEDGEMENTS

We thank Shymala Nada-Raja who provided helpful comments on preliminary work. R.Y., PW and H.S. are supported by the MRC. M.v.B. is funded by NHS Lanarkshire and has an honorary attachment to the MRC unit in Glasgow.

REFERENCES


The Depression Scale as a screening instrument for a subsequent depressive episode in primary healthcare patients

OUTI POUTANEN, ANNA-MAIJA KOIVISTO, MATTI JOUKAMAA, AINO MATTILA and RAIMO K. R. SALOKANGAS

Background There are numerous instruments for screening for depression. A feasible screen is good at both recognising and predicting depression.

Aims To study the ability of the Depression Scale and its items to recognise and predict a depressive episode.

Method A sample of patients attending primary care was examined in 1991–1992 and again 7 years later. The accuracy of the Depression Scale at baseline and at follow-up was tested against the Short Form of the Composite International Diagnostic Interview (CIDI–SF) diagnosis of depression at follow-up. The sensitivity and specificity of the Depression Scale and its items were assessed.

Results Both baseline and follow-up Depression Scale scores were consistent with the CIDI–SF diagnoses. It was possible to find single items efficient at both recognising and predicting depression.

Conclusions The Depression Scale is a useful screening instrument for depression, with both diagnostic and predictive validity.

Declaration of interest None. Funding from the Medical Research Fund of Tampere University Hospital.

There are several instruments to help primary care clinicians identify patients with major depression (Williams et al, 2002). The Depression Scale (Salokangas et al, 1995) is one of these. The relatively low prevalence of depression in primary care practice requires that the sensitivity and specificity of a screening instrument should be almost perfect (Schwenk, 1996). The Beck Depression Inventory (BDI; Beck et al, 1961) and the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snith, 1983) are the most commonly used screening instruments. The popularity of a scale does not guarantee that it is feasible and up-to-date (Bagby et al, 2004). In this study, we aimed to examine the ability of the Depression Scale and its items to recognise and predict a depressive episode.

METHOD

This study forms a part of the larger Tampere Depression Project (TADEP), the baseline study of which was done in 1991–1992 (Salokangas et al, 1995, 1996; Salokangas & Poutanen, 1998). Consecutive patients aged 18–64 years attending primary care services (including consultations in normal office hours and out of hours, occupational health services and visits to prenatal clinics) completed a postal questionnaire including questions on their demographic characteristics, health and functioning, as well as a screening instrument for depression (the Depression Scale; Salokangas et al, 1995). Of the 1643 patients who returned the screening questionnaire adequately filled in, all who screened positive for depression (n=372) and every tenth person who was screen-negative (127 out of 1271 individuals) were invited for interview. To diagnose clinical depression, the Present State Examination (PSE; Wing et al, 1974) was used. A total of 436 persons were interviewed. Their PSE diagnoses were as follows: severe depression n=55, depressive symptoms n=60, other psychiatric symptoms n=174, other psychiatric diagnosis n=29, no psychiatric symptom n=55.

Seven years later a follow-up study was conducted. The number of participants to whom the follow-up questionnaire could be posted was 413 (11 people were dead, no address could be found for 6 and 6 others had attended psychiatric outpatient care and were excluded from subsequent analysis in the present primary care study). Of these 299 returned the questionnaire, and 250 (57.3% of the baseline sample) were willing to take part in the telephone interview. Men (P=0.050) and married individuals (P=0.018) participated more frequently than women or those who were not married. The study protocol was approved by the Tampere University Hospital ethics committee and written informed consent was obtained from the participants.

Study procedure

The Depression Scale includes ten items, with four response alternatives scoring 0–3: ‘not at all’, ‘a little’, ‘quite a lot’ and ‘extremely’ (see Table 2). In the baseline study the cut-off point for the screening sum score was >8.

In the follow-up study participants again filled in the Depression Scale, the Michigan Alcoholism Screening Test (Selzer, 1971), parts of the Hopkins Symptom Checklist (Derogatis et al, 1974), and structured questions. To assess major depressive episode, 38 items from the Short Form of the Composite International Diagnostic Interview (CIDI–SF; World Health Organization, 1989; Kessler et al, 1998) were used in a telephone interview. The CIDI–SF questions concerning the occurrence of symptoms of a major depressive episode referred to the previous month. Three trained psychiatrists (A.M. and Drs Liisa Groth and Niko Seppälä), each with at least 5 years’ experience in psychiatry, conducted the interviews, masked to the baseline PSE diagnoses.

Statistical methods

The accuracy of the Depression Scale as a screening instrument for depression was assessed by receiver operating characteristic (ROC) curve analyses. The follow-up Depression Scale score (DEPS–F) was compared with the CIDI–SF diagnosis of depression. The ability of the baseline Depression Scale score (DEPS–B) to predict
the CIDI–SF diagnosis at follow-up was also evaluated. In ROC analyses, sensitivity, specificity, and areas under the curve were calculated. Sensitivity and specificity were calculated for each reasonable cut-off point of the Depression Scale.

To evaluate which single items of the DEPS–B were best at predicting a depressive episode, the sensitivity and specificity for single items were calculated. After that, logistic regression analysis with forward stepwise method using all DEPS–B items as predictors was conducted. For this analysis, all items were dichotomised using 1 as the cut-off score (0–1, negative item result; 2–3, positive item result). To evaluate which single items of the DEPS–F were best for recognising a depressive episode, the sensitivity and specificity were calculated separately for each item, and logistic regression analysis was likewise conducted.

To identify an ideal pair of Depression Scale items for composing a short version of both DEPS–B and DEPS–F, sensitivity and specificity for every possible DEPS–B and DEPS–F item pair were calculated. An ideal pair of items implied that both of the items scored above 1. Only pairs in which sensitivity was at least 50% were regarded as relevant and reported.

Analyses were performed using the Statistical Package for the Social Sciences version 11.5 for Windows; P<0.05 was considered statistically significant.

**RESULTS**

**Depression Scale v. CIDI–SF**

In participants with CIDI–SF depression, the median DEPS–F score was 18 (range 7–30) and in those without depression it was 5 (range 0–28) (P<0.001, Mann–Whitney test). In the ROC analysis of DEPS–F v. CIDI–SF the area under the curve was 0.939 (Fig. 1). The ideal pair of sensitivity (90.5%, 95% CI 0.71–0.97) and specificity (86.8%, 95% CI 0.82–0.91) was found with a score of >11 as the cut-off point (Table 1). In participants with CIDI–SF depression the median DEPS–B score was 17 (range 2–24) and in those without depression it was 10 (range 0–28) (P<0.001, Mann–Whitney test). In the ROC analysis of DEPS–B v. CIDI–SF the area under the curve was 0.803 (Fig. 1). The ideal pair of sensitivity (72.7%, 92.2%)

### Table 1: Sensitivity and specificity of different Depression Scale cut-off points

<table>
<thead>
<tr>
<th>Depression scale score</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score at follow-up v. CIDI–SF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>95.2</td>
<td>74.4</td>
</tr>
<tr>
<td>9</td>
<td>95.2</td>
<td>78.5</td>
</tr>
<tr>
<td>10</td>
<td>90.5</td>
<td>83.6</td>
</tr>
<tr>
<td>11</td>
<td>90.5</td>
<td>86.8</td>
</tr>
<tr>
<td>12</td>
<td>81.0</td>
<td>89.5</td>
</tr>
<tr>
<td>13</td>
<td>76.2</td>
<td>91.3</td>
</tr>
<tr>
<td>14</td>
<td>71.4</td>
<td>92.2</td>
</tr>
<tr>
<td>15</td>
<td>71.4</td>
<td>94.1</td>
</tr>
</tbody>
</table>

| Score at baseline v. CIDI–SF |               |                 |
| 8                      | 95.5           | 27.3            |
| 9                      | 90.9           | 41.2            |
| 10                     | 86.4           | 53.7            |
| 11                     | 86.4           | 62.5            |
| 12                     | 72.7           | 71.8            |
| 13                     | 72.7           | 77.8            |
| 14                     | 63.6           | 83.8            |
| 15                     | 59.1           | 87.5            |

CIDI–SF, Composite International Diagnostic Interview–Short Form. DEPS, Depression Scale.

**Table 2: Sensitivity and specificity of Depression Scale items at baseline and at follow-up compared with depression assessment with the Composite International Diagnostic Interview.**

<table>
<thead>
<tr>
<th>Depression Scale items¹</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>During the past month I have...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. . . suffered from insomnia</td>
<td>63.6</td>
<td>84.4</td>
<td>54.5</td>
<td>80.0</td>
</tr>
<tr>
<td>2. . . felt blue</td>
<td>59.1</td>
<td>89.3</td>
<td>72.7</td>
<td>74.8</td>
</tr>
<tr>
<td>3. . . felt everything was an effort</td>
<td>81.8</td>
<td>86.4</td>
<td>86.4</td>
<td>63.3</td>
</tr>
<tr>
<td>4. . . felt low energy or slowed down</td>
<td>72.7</td>
<td>83.8</td>
<td>59.1</td>
<td>66.4</td>
</tr>
<tr>
<td>5. . . felt lonely</td>
<td>59.1</td>
<td>93.9</td>
<td>22.7</td>
<td>81.9</td>
</tr>
<tr>
<td>6. . . felt hopeless about the future</td>
<td>81.8</td>
<td>92.5</td>
<td>59.1</td>
<td>74.7</td>
</tr>
<tr>
<td>7. . . not got any fun out of life</td>
<td>54.5</td>
<td>84.8</td>
<td>27.3</td>
<td>73.5</td>
</tr>
<tr>
<td>8. . . had feelings of worthlessness</td>
<td>50.0</td>
<td>96.1</td>
<td>45.5</td>
<td>81.0</td>
</tr>
<tr>
<td>9. . . felt all pleasure and joy has gone from life</td>
<td>45.5</td>
<td>93.8</td>
<td>59.1</td>
<td>82.8</td>
</tr>
<tr>
<td>10. . . felt that I cannot shake off the blues even with help from family and friends</td>
<td>42.9</td>
<td>91.2</td>
<td>36.4</td>
<td>81.0</td>
</tr>
</tbody>
</table>

¹All items are scored 0, not at all; 1, a little; 2, quite a lot; 3, extremely. An item was regarded as positive when the score was >1.
Table 3  Depression Scale items at baseline and at follow-up from logistic regression analyses significantly associated with depression at follow-up assessment.

<table>
<thead>
<tr>
<th>Depression Scale items</th>
<th>OR (95 % CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPS at follow-up v. depression at follow-up (CIDI–SF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>During the past month I have . . .</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 . . . felt everything was an effort</td>
<td>5.54</td>
<td>(1.35–22.79)</td>
</tr>
<tr>
<td>6 . . . felt hopeless about the future</td>
<td>21.89</td>
<td>(5.45–88.01)</td>
</tr>
<tr>
<td>DEPS at baseline v. depression at follow-up (CIDI–SF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>During the past month I have . . .</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 . . . suffered from insomnia</td>
<td>2.67</td>
<td>(0.99–7.19)</td>
</tr>
<tr>
<td>3 . . . felt everything was an effort</td>
<td>6.50</td>
<td>(1.76–24.01)</td>
</tr>
<tr>
<td>9 . . . felt all pleasure and joy has gone from life</td>
<td>3.70</td>
<td>(1.35–10.09)</td>
</tr>
</tbody>
</table>

CIDI–SF, Composite International Diagnostic Interview–Short Form; DEPS, Depression Scale.

Table 4  Sensitivity and specificity of Depression Scale item pairs at baseline and at follow-up compared with depression at follow-up assessment.

<table>
<thead>
<tr>
<th>DEPS score v. CIDI–SF episode of depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPS at follow-up</td>
</tr>
<tr>
<td>DEPS at baseline</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>Specificity (%)</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>Specificity (%)</td>
</tr>
<tr>
<td>DEPS item pair</td>
</tr>
<tr>
<td>1+3</td>
</tr>
<tr>
<td>59.1</td>
</tr>
<tr>
<td>45.5</td>
</tr>
<tr>
<td>2+3</td>
</tr>
<tr>
<td>50.0</td>
</tr>
<tr>
<td>72.7</td>
</tr>
<tr>
<td>2+6</td>
</tr>
<tr>
<td>59.1</td>
</tr>
<tr>
<td>50.0</td>
</tr>
<tr>
<td>2+9</td>
</tr>
<tr>
<td>27.3</td>
</tr>
<tr>
<td>54.5</td>
</tr>
<tr>
<td>3+4</td>
</tr>
<tr>
<td>72.7</td>
</tr>
<tr>
<td>59.1</td>
</tr>
<tr>
<td>3+5</td>
</tr>
<tr>
<td>54.5</td>
</tr>
<tr>
<td>22.7</td>
</tr>
<tr>
<td>3+6</td>
</tr>
<tr>
<td>72.7</td>
</tr>
<tr>
<td>54.5</td>
</tr>
<tr>
<td>3+9</td>
</tr>
<tr>
<td>45.5</td>
</tr>
<tr>
<td>59.1</td>
</tr>
<tr>
<td>4+6</td>
</tr>
<tr>
<td>68.2</td>
</tr>
<tr>
<td>36.4</td>
</tr>
<tr>
<td>5+6</td>
</tr>
<tr>
<td>54.5</td>
</tr>
<tr>
<td>13.6</td>
</tr>
</tbody>
</table>

Depression Scale items v. CIDI–SF

The three most sensitive DEPS–F items were 3 (‘I have felt everything was an effort’), 6 (‘I have felt hopeless about the future’) and 4 (‘I have felt low energy or slowed down’), and the most specific items were 8 (‘I have had feelings of worthlessness’), 5 (‘I have felt lonely’) and 9 (‘I have felt all pleasure and joy has gone from life’) (Table 2). In the case of DEPS–B, item 3 had a high sensitivity whereas items 9, 5, 8 and 10 (‘I felt that I cannot shake off the blues even with help from family and friends’) had a reasonably high specificity. One item (item 3) was quite sensitive in both analyses, for both recognising and predicting CIDI–SF depression.

In logistic regression analyses, DEPS–F items 3 and 6 were significantly associated with CIDI–SF depression, whereas DEPS–B items 1 (‘I have suffered from insomnia’), 3 and 9 significantly predicted occurrence of subsequent CIDI–SF depression (Table 3).

Best Depression Scale item pairs v. CIDI–SF

Sensitivity and specificity were calculated for every possible pair of Depression Scale items to ascertain which two items had the best balance of recognition and prediction. Only the pairs with sensitivity of at least 50% are reported (Table 4). The three best pairs for recognition were items 3 and 6, items 3 and 4, and items 4 and 6, whereas the best pairs for prediction were items 2 (‘I have felt blue’) and 3, items 3 and 4, and items 3 and 9.

DISCUSSION

The Depression Scale was quite consistent with the CIDI–SF both as a predictor and a recogniser of depression. ‘Feeling that everything is an effort’ and ‘feeling hopeless about the future’ were the best items, and also the best item pair for recognising depression. ‘Suffering from insomnia’ ‘feeling everything is an effort’ and ‘feeling all pleasure and joy were gone from life’ were the best items for predicting future depression. ‘Feeling blue’ and ‘feeling everything is an effort’ were the best item pair for predicting future depression.

Sensitivity and specificity

The first validation of the Depression Scale was reported in an earlier study, in which the cut-off point for depression was >8 (Salokangas et al., 1995). In the baseline validation study, using the PSE as the criterion, the sensitivity of the Depression Scale for clinical depression was 74% and the specificity for non-depression 85%. For severe depression the figures were 84% and 93%. In the present study the figures for sensitivity and specificity were better than those of the earlier validation study. In the baseline validating analyses the sampling ratio was taken into account, but this was not done in the present study, which was mainly intended to ascertain the ability of the scale to predict an episode of depression and to evaluate its individual items. The differences in the levels of sensitivity and specificity between the baseline validation analyses and these follow-up analyses are perhaps partly explained by this fact. There are also differences in the validity criterion between the two diagnostic instruments. The PSE is based on symptoms, and the CIDI is based on syndromes (Lowe et al., 2004). With the CIDI–SF the definition of depression was clearer because there were only two categories: depressive and non-depressive. It should also be kept in mind that the PSE interviews at baseline were held face-to-face, whereas the CIDI–SF interviews at follow-up were...
conducted by telephone. A telephone interview relies more on the examinee’s own assessment, and is closer to a self-rating instrument like the Depression Scale. The same items of the CIDI-SF were used as in a previous Finnish depression study (Isometsa et al., 1997; Lindeman et al., 2000) using the computer-assisted telephone interview method.

According to Lowe et al. (2004) the sensitivity of screening questionnaires should lie above specificity and be as high as possible, and the specificity should be at least 75%. In this study the cut-off point >11, which has a sensitivity of 90.5% and specificity of 86.8%, could be ideal.

When the ability of the Depression Scale to predict an episode of depression was analysed, the area under the curve was 0.803. An earlier study with primary care patients (Salokangas et al., 1994) showed that the rate of clinical depression in people with a Depression Scale score above 12 was about 47% and in those with a score above 15 it was about 57%. These percentages are high enough to have some clinical value. In this study, with a cut-off point of >11 sensitivity was 86.4% but specificity only 62.5%. When an instrument is used as a predictor it is perhaps more important to avoid false positives and not to stigmatise patients; this justifies a higher cut-off point.

**What did the Depression Scale actually assess?**

In a study of general practice patients (Williamson et al., 2005), four mental health self-report scales and a composite of those four were assessed to determine their accuracy in predicting psychiatric caseness for depression, dysthymia, generalised anxiety disorder, social phobia, agoraphobia and panic attack. One scale measuring neuroticism – the Neuroticism Scale of the Eysenck Personality Questionnaire (EPQ-N; Eysenck et al., 1985) – and a composite of all four scales were found to be very strong and accurate predictors of psychiatric caseness, but they were unable to differentiate between specific disorders. In our study only episode of depression – not other psychiatric diagnoses – was assessed.

In an extensive follow-up study (Tyrer et al., 2004) the quick-to-use HADS was good for recognising both depression and anxiety, and was better than any other single measure for predicting the outcome of both anxiety and depressive disorders after an interval of 12 years. The Montgomery–Asberg Depression Rating Scale did not have such predictability.

When the Depression Scale and two common self-rating instruments (the BDI and the HADS) are compared, they differ in many ways. The Depression Scale concentrates on the previous month, whereas the BDI concentrates on the previous week (the BDI-II on the past 2 weeks; Beck et al., 1996) and the HADS on current feelings. Of the criterion standards used in both, the PSE and the CIDI-SF refer to the previous month. It is difficult to say, however, what the true significance of the differences in these time periods is.

The Depression Scale is the shortest of the three instruments, and the BDI is the longest. The formulation of the items is different: the most evident difference is that the Depression Scale gives exactly the same short-answer alternatives for all ten items, whereas there are several different sets of alternative answers in both the BDI and the HADS. This makes the Depression Scale very quick and easy to use, and increases adherence.

The BDI includes most of the Depression Scale topics. Only the topics of items 5 (loneliness), 7 (no fun) and 10 (not helped even with family and friends) are missing in the BDI. The Depression Scale item 5 was specific in recognising depression and item 10 specific in predicting it. However, the BDI covers the symptoms of depression more comprehensively than the former scale. The HADS covers both depression and anxiety, but lacks most of the Depression Scale topics (items 1, 2, 3, 5, 8 and 10); the symptoms covered are less severe than in the BDI or in the Depression Scale. Common topics for all the three self-rating instruments are the Depression Scale items 4 (low energy), 6 (hopelessness), and 9 (lost pleasure and joy). These topics probably relate to the core of depression symptomatology; other topics can be said to be consequences of the core symptoms and not so essential to depression only.

The Depression Scale items 3 and 4 were good at both recognising and predicting depression. Item 3 (‘I have felt everything was an effort’) suggests reduction of energy, which is one of the main symptoms of depression according to the ICD–10. Item 6 was good for recognition even though its wording refers to the future (‘I have felt hopeless about the future’); hopelessness is also a symptom of depression in the ICD–10. Item 9 was good in predicting depression. The wording of item 9 (‘I have felt all pleasure and joy has gone from life’) refers to something that has already happened, something that is possibly endured as beyond help. Item pair 2 and 3 was the best at predicting depression. The wording of item 2 (‘I have felt blue’) may be experienced as persistent low mood, referring to a more chronic state. It is almost the same as lowering of mood, one of the main symptoms of depression in ICD–10. The best combination – and a possible quick version – of two items for recognising depression was items 3 and 6, and the best combination for predicting depression was items 2 and 3.

The use of psychometric scales is in general problematic. Among people who appear to be healthy according to standard mental health scales it is possible to identify a subgroup of people who may not be psychologically healthy at all: mental health scales may assess not mental health but instead defensive denial (Shedler et al., 1993). Moreover, any scale that is valid for assessing current depression will have some long-term predictability because depression is recurrent. However, if a scale has predictability, it means it has the ability to catch not just reactive and short-term symptoms but more chronic or recurrent core features of the disorder.

**Limitations and strengths of the study**

It is a limitation of the study that the interviews were held by telephone. However, the CIDI–SF telephone interviews were conducted with care and by experienced psychiatrists. Some information about the mental state of these patients during the follow-up period was gathered, but this was self-report information and possibly not so reliable, and we decided not to use it in this study. This was not a follow-up study in its truest sense: the assessments were made only twice – at baseline and 7 years later. Thus, the mental state of the participants during the intervening period is obscure, decreasing slightly the credibility of the study. It is strength of the study that the sample was fairly large, and that it was a follow-up study with a wide range of primary care patients.

**Implications**

The Depression Scale is not only an easy-to-use screening instrument, it also appears to
The 12-month prevalence and risk factors for depression in Finland: a computer assisted telephone interview study.


Heroin-assisted treatment for opioid dependence

Randomised controlled trial

CHRISTIAN HAASEN, UWE VERTHEIN, PETER DEGKWITZ, JUERGEN BERGER, MICHAEL KRAUSZ and DIETER NABER

Background  Heroin-assisted treatment has been found to be effective for people with severe opioid dependence who are not interested in or do poorly on methadone maintenance.

Aims  To study heroin-assisted treatment in people on methadone who continue intravenous heroin and in those who are heroin dependent but currently not in treatment.

Method  In an open-label multicentre randomised controlled trial, 1015 people with heroin dependence received a variable dose of injectable heroin (n=515) or oral methadone (n=500) for 12 months. Two response criteria, improvement of physical and/or mental health and decrease in illicit drug use, were evaluated in an intent-to-treat analysis.

Results  Retention was higher in the heroin (67.2%) than in the methadone group (40.0%) and the heroin group showed a significantly greater response on both primary outcome measures. More serious adverse events were found in the heroin group, and were mainly associated with intravenous use.

Conclusions  Heroin-assisted treatment is more effective for people with opioid dependence who continue intravenous heroin while on methadone maintenance or who are not enrolled in treatment. Despite a higher risk, it should be considered for treatment resistance under medical supervision.

Declaration of interest  None.

Funding detailed in Acknowledgements.

Germany has an estimated 150,000 people with opioid dependence, mainly heroin dependence, among a population of 80 million (Buhringer et al, 1997). Less than half (50,000–60,000) at any given time are on opioid maintenance treatment. None the less, the mortality rate only decreased slightly after the widespread introduction of maintenance treatment in the early 1990s (Raschke et al, 2000), which is in accordance with other long-term follow-up studies (Rathod et al, 2005). This opened the discussion for modification of maintenance treatment, especially for people who either dropped out or who continued treatment but also illicit opioid use.

A large (n=1969) cohort study was initiated in Switzerland in 1994, and ascertained the feasibility, safety and potential efficacy of offering injectable heroin to people with dependence who were not responding sufficiently to maintenance treatment (Rehn et al, 2001). The study showed a high retention rate (70% after 12 months) as well as positive effects with respect to illegal drug use, physical and mental health and social outcomes. However, assessment of the Swiss trial by the World Health Organization was unable to determine if the positive effects were a result of the prescription of heroin, the extensive psychosocial counselling and care, or the combination of both (Ali et al, 1999).

A small randomised controlled trial (n=51) comparing injectable heroin with a standard treatment (mainly methadone maintenance) showed significantly better functioning in those receiving heroin after 6 months (Perneger et al, 1998). However, those people also received additional, mandatory psychosocial care, which may have influenced the results.

In 1998 two randomised controlled trials in The Netherlands assessed the effectiveness of the co-prescription of inhalable (n=375) and injectable (n=174) heroin in people with opioid dependence and chronic resistance to methadone treatment. Results showed that heroin-assisted treatment was feasible, more effective and probably as safe as methadone alone in reducing physical, mental and social problems (van den Brink et al, 2003; Blanken et al, 2005).

Co-prescription of heroin was cost-effective compared with methadone treatment alone (Dijkgraaf et al, 2005). A limitation of these trials was that psychosocial treatments were not standardised and were uncontrolled. Furthermore, the larger of the two trials used inhalable heroin, which is used by the majority (75–90%) of street heroin users in The Netherlands, but not in Germany.

A recent Cochrane review (Ferri et al, 2005) found that the Swiss and Dutch studies do not allow a definite conclusion to be drawn about the overall effectiveness of heroin prescription because of a lack of comparability. We therefore examined the effectiveness of medically prescribed and supervised heroin injection in an open-label randomised controlled trial in two groups of people with heroin dependence: those not responding sufficiently to methadone maintenance treatment and those currently not in substance misuse treatment. To control for the impact of psychosocial treatment, participants in each group were randomised to one of two types of psychosocial care.

METHOD

Study design

After screening more than 2000 people with heroin dependence, a total of 1032 consenting participants were randomised between March 2002 and December 2003 in seven treatment centres (Hamburg, 401 participants; Frankfurt, 191; Hanover, 132; Bonn, 100; Cologne, 100; Munich, 60; Karlsruhe, 48). Participants were from two target groups: (a) people with heroin dependence who were insufficiently responding to treatment owing to continuous intravenous heroin use (n=492); and (b) people with heroin dependence who were not in treatment in the previous 6 months (n=540). Participants from each target group were randomised into four subgroups according to the type of medication and the type of psychosocial care (Fig. 1), resulting in a 2 x 2 x 2 design and eight separate groups. Of the 811 people lost between screening and baseline, 106 (13.1%) did not meet inclusion criteria and the
CONSORT diagram. ITT, intent-to-treat.

56

sent after randomisation without initiating analysis because they withdrew their consent. Seventeen patients, 5 previously on methadone and 12 not in treatment, were excluded from analysis because they withdrew their consent after randomisation without initiating analysis (prior to randomisation (n=8), because they did not have an independent baseline interview prior to randomisation (n=8), or both (n=1), leaving 1015 patients in the intent-to-treat analysis (n=487 treatment failure, n=528 not in treatment).

After giving consent, participants were given an extensive baseline examination. Inclusion criteria were then presented to a local independent expert committee before a final decision for inclusion was made. Then a second consent was necessary before randomisation. Randomisation took place separately for each target group (methadone treatment failure and not in treatment), and treatment allocation was performed using sealed and consecutively numbered envelopes at each study site.

Treatment duration was 12 months. Treatment in the intervention group consisted of an individually adjusted dose of injectable heroin that was self-administered in an out-patient setting under direct supervision of medical staff, maximally three times a day, 7 days a week, with a maximum single dose of 400 mg and a maximum daily dose of 1000 mg (none to take home). Up to 60 mg of methadone could also be given for take-home night-time use to suppress withdrawal. Treatment in the control group consisted of a minimum daily dose of 60 mg methadone, which could be individually adjusted according to clinical judgement. Participants within both groups were randomised to either group psychoeducation plus individual counselling according to Farnbacher et al (2002), or case management and motivational interviewing according to Oliva et al (2001). Each of these interventions has been described in manuals, and training of all therapists was conducted prior to the study to minimise site differences. The type of psychosocial care was similar with respect to average intensity of contact, but there was more individual flexibility in the case management group than with the more standardised psychosocial care in the psychoeducation group.

Study population

Inclusion criteria were 23 years old or greater and an ICD-10 diagnosis of opioid dependence of at least 5 years’ duration (World Health Organization, 1993). Furthermore, eligibility criteria for the group with methadone treatment failure included continued intravenous use of street heroin (confirmed by urine testing) despite ongoing maintenance treatment of at least 6 months, whereas for the not in treatment group they included regular intravenous use of street heroin (confirmed by urine testing) and confirmed participation in previous drug treatment. Participants needed to have poor physical and/or mental health, with at least 13 symptoms on the Opiate Treatment Index (OTI) Health Scale (Darke et al, 1991, 1992) and/or at least 60 points (standardised T-score) on the Global Severity Index of the Symptom Check-List (SCL-90-R; Derogatis, 1983).

People with a pending jail sentence, those who had been abstinent for 2 or more months in the past 12 months and those
with a severe physical disorder such as renal or hepatic failure, clinically significant cardiac arrhythmias or chronic obstructive pulmonary disease were excluded, as were pregnant or breast-feeding women.

Assessments and statistical analyses

Baseline assessments were completed by study physicians and independent research assistants before a decision was made on randomisation. Potential study inclusion was based on physician assessment only but had to be confirmed by an independent panel of experts after baseline assessment, which delayed initiation of treatment for an average of 31 days. Study physicians re-assessed people who were approved for randomisation at initiation of treatment, and at 1, 3, 6 and 12 months. Independent assessment by research assistants was performed at 6 and 12 months.

Assessment by the study physician included application of the OTI and SCL–90–R, the composite international diagnostic interview (CIDI; World Health Organization, 1990), and the severity of withdrawal scale (SOWS; Gossop, 1990), and a comprehensive physical examination, including electrocardiography, laboratory examinations, echocardiography, abdominal ultrasonography, urine and hair analyses, as well as all serious adverse events. All serious adverse events, defined according to guidelines E2A and E6 of the International Conference on Harmonisation of Technical Registration for Recognition of Pharmaceuticals for Human Use (ICH; http://www.ich.org) were reported to a safety board, which consisted of three independent clinicians, who evaluated all adverse events with respect to safety of the study treatment. The assessment by independent research assistants included administration of the European version of the Addiction Severity Index (EuropASI; Kokkevi & Hartgers, 1995), and gathering data on criminal behaviour and on subjective aspects of treatment. In the intent-to-treat analysis, all those randomised were assessed regardless of treatment retention. Data from the baseline and 12-month assessments were used for analysis of the primary outcome measures; the last-observation-carried-forward (LOCF) procedure from data at 6 months was used if data at 12 months were missing. If no data were available for 6 and 12 months, the outcome was coded according to a worst-case analysis (i.e. as a responder in the methadone group and a non-responder in the heroin group).

Two prespecified dichotomous, multi-domain primary outcome measures were used. For the primary outcome measure on health, participants were considered responders if they showed at least a 20% improvement and at least 4 points on the OTI Health Scale (physical health) and/or at least a 20% improvement in the GSI (mental health), without a deterioration of more than 20% in the other area of health. For the second primary outcome measure, people were considered responders if they showed a reduction in the use of street heroin with at least 3 of 5 urine samples negative for the drug in the month prior to the 12-month assessment and no increase in cocaine use (hair analysis). If less than 3 urine samples or no hair was available at 12 months, data from urine or hair testing at 6 months were used (LOCF). If these were also not available, data were replaced by self-reported data from the EuropASI. When self-reported data were used, response was defined as a 60% decrease in the number of days with street heroin use and no more than 2 days’ increase in cocaine use during the past month. To distinguish between prescribed and illicit heroin, urine samples were tested for papaverine and acetylcocaine, which are common impurities found in street heroin (Paterson et al., 2005; Rook et al., 2006).

A four-factorial logistic regression model was used to assess the effectiveness of heroin-assisted treatment compared with methadone, controlling for the effect of the target group (methadone treatment failure vs. not in treatment), the psychosocial intervention (psychoeducation vs. case management) and study site (likelihood ratio test). Using a test on interaction between primary outcome and target group (methadone treatment failure or not in treatment), we assessed whether the effect of pharmacological treatment was independent of the target group. The hypothesis would be confirmed if the logistic regression model showed superiority of heroin over methadone for both primary outcome measures (‘health’ and ‘illegal drug use’) at the 5% significance level. Statistical analyses were performed using SPSS versions 10 and 11 for Windows.

Calculations of sample size were based on an estimated response rate of 30% in the methadone group and 50% in the heroin group for each primary outcome measure. Based on a one-tailed significance criterion of 0.025 (α) and a β of 0.90 for each primary outcome measure, the total power remained 80% (0.9*0.9) for the study to yield a statistically significant result. Assuming that 10% of the methadone group and 5% of the heroin group would not be reached for assessment at 6 or 12 months, and therefore according to the worst case definition would be considered responders and non-responders respectively, the reduced effect size led to a minimum sample size of 482 for each treatment group (heroin vs. methadone).

RESULTS

Sample characteristics

Table 1 shows baseline characteristics of the participants included in the intent-to-treat analysis. Both target groups had severe drug use, health problems and social problems. The group not in treatment had a more severe pattern of drug use and more problems with housing than those with past methadone treatment failure. Of the 487 in the treatment failure group, 387 were previously being treated with methadone (mean dose 90.6 mg/day), 64 with levomethadone (mean dose 56.4 mg/day), 33 with buprenorphine (mean dose 10.7 mg/day), and 3 with dihydromorphine (mean dose 2080.0 mg/day).

Availability of outcome data

Follow-up data were available at 12 months for 956 of the 1015 participants (95.1% of the heroin group and 93.2% of the methadone group). Health data were available for 970 patients (497 from the heroin group and 473 from the methadone group, including LOCF and death cases), leaving 45 instances where missing response data had to be replaced according to the worst case strategy. Data on illicit drug use were available for 982 participants (504 from the heroin group and 478 from the methadone group, including LOCF and death cases), leaving 33 instances where missing response data had to be replaced according to the worst case strategy.

Treatment retention

Treatment retention was higher in the heroin group, with 67.2% completing 12-month treatment compared with 40.0% in the methadone group. However, 28.8% of the methadone group did not even initiate
Table 1  Baseline characteristics of 1015 people with heroin dependence who participated in the study

<table>
<thead>
<tr>
<th></th>
<th>Methadone treatment failure</th>
<th></th>
<th>Not in treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heroin</td>
<td>Methadone</td>
<td>Total</td>
</tr>
<tr>
<td>Male gender, %</td>
<td>78.5</td>
<td>77.2</td>
<td>77.8</td>
</tr>
<tr>
<td>Age, years: mean (s.d.)</td>
<td>36.7 (6.5)</td>
<td>37.1 (6.7)</td>
<td>36.9 (6.6)</td>
</tr>
<tr>
<td>Stable housing, %</td>
<td>74.8</td>
<td>75.5</td>
<td>75.2</td>
</tr>
<tr>
<td>Employed, %</td>
<td>6.1</td>
<td>3.7</td>
<td>4.9</td>
</tr>
<tr>
<td>Regular drug use, years: mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heroin</td>
<td>14.2 (6.2)</td>
<td>14.4 (6.3)</td>
<td>14.3 (6.3)</td>
</tr>
<tr>
<td>Cocaine</td>
<td>6.1 (6.9)</td>
<td>5.9 (6.4)</td>
<td>6.0 (6.7)</td>
</tr>
<tr>
<td>Benzodiazepines</td>
<td>6.2 (7.8)</td>
<td>7.3 (7.8)</td>
<td>6.7 (7.8)</td>
</tr>
<tr>
<td>Drug use in past month, days: mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heroin</td>
<td>17.1 (10.8)</td>
<td>17.6 (10.5)</td>
<td>17.4 (10.7)</td>
</tr>
<tr>
<td>Cocaine</td>
<td>14.7 (10.0)</td>
<td>14.1 (10.8)</td>
<td>14.4 (10.9)</td>
</tr>
<tr>
<td>Benzodiazepines</td>
<td>18.7 (11.2)</td>
<td>18.4 (11.5)</td>
<td>18.6 (11.3)</td>
</tr>
<tr>
<td>Intravenous drug use</td>
<td>19.7 (10.7)</td>
<td>20.3 (10.5)</td>
<td>20.0 (10.6)</td>
</tr>
<tr>
<td>Alcohol use in past month, days: mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous detoxification treatment, %</td>
<td>88.1</td>
<td>90.4</td>
<td>89.2</td>
</tr>
<tr>
<td>Previous drug-free treatment, %</td>
<td>62.6</td>
<td>61.1</td>
<td>61.8</td>
</tr>
<tr>
<td>Previous maintenance treatment, %</td>
<td>100.0</td>
<td>99.6</td>
<td>99.8</td>
</tr>
<tr>
<td>Physical health score, mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTI Health Scale score, mean (s.d.)</td>
<td>18.8 (5.1)</td>
<td>18.9 (5.5)</td>
<td>18.9 (5.3)</td>
</tr>
<tr>
<td>Body mass index</td>
<td>23.0 (3.8)</td>
<td>22.9 (3.8)</td>
<td>22.9 (3.8)</td>
</tr>
<tr>
<td>HIV positive, %</td>
<td>11.8</td>
<td>10.9</td>
<td>11.4</td>
</tr>
<tr>
<td>HCV positive, %</td>
<td>82.8</td>
<td>85.4</td>
<td>84.1</td>
</tr>
<tr>
<td>Mental health</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSI standardised T-score: mean (s.d.)</td>
<td>69.5 (11.0)</td>
<td>69.7 (9.8)</td>
<td>69.6 (10.4)</td>
</tr>
<tr>
<td>Previous suicide attempts, %</td>
<td>45.8</td>
<td>43.5</td>
<td>44.6</td>
</tr>
<tr>
<td>At least one lifetime psychiatric diagnosis, %</td>
<td>62.1</td>
<td>60.8</td>
<td>61.7</td>
</tr>
<tr>
<td>Social functioning score</td>
<td>GAFS: mean (s.d.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>53.3 (10.5)</td>
<td>52.5 (11.9)</td>
<td>52.9 (11.2)</td>
</tr>
<tr>
<td>Illegal activities past month, days, mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18.8 (11.0)</td>
<td>18.8 (10.5)</td>
<td>18.8 (10.7)</td>
</tr>
<tr>
<td>Ever convicted, %</td>
<td>97.1</td>
<td>96.2</td>
<td>96.6</td>
</tr>
<tr>
<td>Ever incarcerated, %</td>
<td>74.2</td>
<td>76.0</td>
<td>75.1</td>
</tr>
</tbody>
</table>

OTI, Opiate Treatment Index; HCV, hepatitis C virus; GSI, Global Severity Index; GAFS, Global Assessment of Functioning Scale.

*P < 0.05 methadone treatment failure v. not in treatment.
1. ICD-10 diagnosis of schizophrenic (F2), affective (F3), neurotic (F4) or behavioural (F5) disorder.
2. One participant did not meet criteria for 6-month maintenance treatment in independent assessment.

study treatment (in contrast to 2.3% of the heroin group). Of those initiating treatment, 68.3% of the heroin group and 56.3% of the methadone group completed study treatment; 11.8% of the heroin group and 24.8% of the methadone group started with an abstinence-based or maintenance treatment after dropping out of the study treatment. The average number of treatment days was 290 days in the heroin group and 195 days in the methadone group. The mean daily dose of heroin was 442 mg with an additional 8 mg of methadone (mean daily dose over all heroin treatment days) – additional methadone was only necessary on 20.6% of heroin treatment days. In the methadone group the mean daily dose was 99 mg methadone.

Treatment effectiveness

In the intent-to-treat analysis, the heroin treatment group showed a significantly greater response than the methadone treatment group with respect to both primary outcome measures (Table 2).

With respect to the primary outcome measure ‘health’, logistic regression analysis showed no effect of target group (methadone treatment failure v. not in treatment; P=0.320), study centre (P=0.143) and type of psychosocial intervention (psycho-education v. case management; P=0.269). In addition, no interaction was found between medication group and target group (P=0.544). After adjustment for target group, study centre and type of psychosocial care, the main effect of medication group on the primary outcome measure ‘health’ remained significant (OR=1.54, 95% CI 1.02–2.34, P=0.042).

With respect to the primary outcome measure ‘illicit drug use’, a significant effect of study centre was found (P=0.002), indicating that response rates were not
homogenous across centres. Target group (P=0.228) and type of psychosocial care (P=0.369) showed no significant effect. Furthermore, no interaction was found between medication effect and target group (P=0.840). After adjustment for target group, study centre and type of psychosocial care, the main effect of medication group on the primary outcome measure ‘illicit drug use’ remained significant (OR=1.91, 95% CI 1.30–2.79, P=0.001).

Of the 1015 patients included in the intent-to-treat analysis, 546 (346 in the heroin group and 200 in the methadone group) completed the study as defined per protocol. In those 546 participants the response rates were slightly higher than in the intent-to-treat analysis, but the heroin group also showed a significantly greater response rate than the methadone group (Table 2).

Using a more conservative analysis strategy that defined responders as only those patients responding on both primary outcome measures, the intent-to-treat analysis showed a significantly greater response rate in the heroin compared with the methadone group (57.3% v. 44.8% OR=1.67, 95% CI 1.30–2.14, P<0.001). Using this strategy analysis of the 546 participants completing the study also showed a significantly better response rate for the heroin than the methadone group (63.6 v. 39.5%, OR=2.73, 95% CI 1.88–3.97, P<0.001).

Physical health (OTI Health Scale) showed a significant improvement in both groups, with the greatest improvement observed during the time while preparing for initiation of treatment and the first month of treatment (Fig. 2). The assessment of illicit drug use (according to self-reported data) showed a marked reduction of street heroin use in both groups, but a more pronounced reduction in the heroin group, and a moderate reduction of cocaine use in both groups (Fig. 3). Urine testing at 6 and 12 months for street heroin, as well as weekly urine testing for cocaine, confirms the self-reported data (Fig. 4). Hair analysis for cocaine use confirmed results of urine testing and self-reported data, showing an overall decrease in cocaine use, but especially a decrease in intensive use (from 29.5 to 17.2% of samples in the heroin group and 31.6 to 22.4% in the methadone group).

### Safety

A total of 315 serious adverse events were reported during the 12-month study period: 177 among 124 participants in the heroin group and 138 among 88 participants in the methadone group (Table 3). In 58 instances (32.8%) in the heroin group, the adverse event was possibly, probably or definitely related to the study medication, whereas in the methadone group this occurred less often (15 serious adverse events, 10.9%).

Of the 58 adverse events possibly, probably or definitely related to the heroin medication, 41 occurred within a few minutes of injection, 31 of these events were related to respiratory depression, in most cases associated with unreported concomitant illicit benzodiazepine use, whereas 10 were related to an epileptic seizure. Considering the longer average length of per-protocol treatment in the heroin compared with the methadone group (149 330 v. 97 500 cumulative treatment days), a serious adverse event that was possibly, probably or definitely related to the study medication occurred 2.5 times more often (every 2572 v. 6501 treatment days in the heroin and methadone groups respectively). There were 12 deaths (5 in heroin group, 7 in methadone group) in the 12-month study period for the intent-to-treat population. Of these only 5 occurred while the participant was using study medication and none were possibly, probably or definitely related to the study medication (3 in heroin group: 1 spleen rupture after falling, 1 intoxication with illicit methadone 1 owing to pneumonia and myocarditis; 2 in methadone group: 1 ruptured aneurysm, 1 reason unknown but no methadone in days before death).

### DISCUSSION

### Main findings

This randomised controlled trial found that heroin-assisted treatment of people with severe opioid dependence and treatment...
Table 3  Serious adverse events in intent-to-treat population during 12-month study period

<table>
<thead>
<tr>
<th></th>
<th>Heroin</th>
<th>Methadone</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Events</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possibly related to heroin or methadone</td>
<td>177</td>
<td>100</td>
<td>138</td>
</tr>
<tr>
<td>Probably/definitely related to heroin or methadone</td>
<td>34</td>
<td>19.2</td>
<td>8</td>
</tr>
<tr>
<td>Possibly/probably/definitely related to heroin or methadone</td>
<td>24</td>
<td>13.6</td>
<td>7</td>
</tr>
<tr>
<td>or methadone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Related to intravenous application</td>
<td>58</td>
<td>32.8</td>
<td>15</td>
</tr>
<tr>
<td>Treatment days until occurrence of event possibly/probably/definitely related to study medication</td>
<td>2572</td>
<td>6501</td>
<td>3382</td>
</tr>
</tbody>
</table>

1. 149,350 cumulative treatment days in heroin group.
2. 97,500 cumulative treatment days in methadone group.

Fig. 3  Change in street heroin and cocaine use in the past 30 days (self-reported data); —— heroin; —— methadone; self-reported data were collected by the attending physician, whenever possible missing values were completed with data from independent interviews.

Fig. 4  Testing of urine samples for street heroin (left) and cocaine (right) during the study period; —— heroin; —— methadone.

indicate that a well-structured treatment with trained therapists using standardised and clinically relevant psychosocial interventions can lead to positive outcomes even in a group that has previously responded poorly to methadone treatment. The confirmation of the positive results in the heroin group in the per-protocol analysis is of importance because a positive outcome in the methadone group was expected owing to a low retention rate (highly selected group) but remained significantly below the positive outcome in the heroin group.

These positive effects of heroin-assisted treatment should be weighted against the higher rate of serious adverse events which appear to be associated with the route of administration of opioids and are not unexpected. However, the controlled clinical setting for heroin treatment, with a required 30 min stay after intravenous injection, allows adverse events to be easily managed clinically, unlike when street heroin is injected in uncontrolled and unhygienic settings. No fatalities occurred that were possibly, probably or definitely related to the study medication in either group. The rate of serious adverse events was higher than in the Dutch study (van den Brink et al., 2003), which may be because in the latter study heroin-assisted treatment was supplementary to methadone maintenance treatment.

This study confirms in a large sample the positive effects of heroin-assisted treatment reported from uncontrolled (Rehm et al., 2001) and controlled (Perneger et al., 1998; van den Brink et al., 2003) trials for people resistant to methadone treatment. These data also show that heroin-assisted treatment can be helpful for those with heroin dependence currently not in treatment. It should be noted, however, that many of the latter group have an extensive treatment history and their baseline characteristics were similar to the methadone patients. The use of two structured psychosocial interventions in each treatment condition suggests that the observed differences between the methadone and heroin groups were not the result of differences in psychosocial treatment.

Another methodological strength of the study is the conservative analysis strategy, using a worst case strategy for all missing data not replaced by LOCF. Considering the nature of this group of patients, the high rate of adherence, with 12-month data for most participants, strengthens the interpretation of the results. Despite a general preference for other methods such as direct likelihood analysis or multiple imputation for missing data, in this study these methods would have reproduced differences in distribution of missing values, whereas the LOCF procedure allowed only data collected after 6 months to replace missing data and mirrors more actual treatment effects. Considering the high drop-out rate in the methadone group, a LOCF.
procedure leads to more results of patients still in treatment, therefore favouring the overall results of the methadone group.

Limitations of the study

Given the nature of the medication under study, a double-blind design was not possible (Bammer et al., 1999). Furthermore, the response rates for the primary outcome measure ‘health’ were much higher for both groups than expected, so that the extent of improvement defined as a response may have been too low. Therefore, a sensitivity analysis using the worst case strategy and a 40% improvement as a definition of response was performed, in order to better compare the results with the Dutch study (van den Brink et al., 2003). This showed that lower response rates were observed, but the response rate for the heroin group remained significantly higher than that for the methadone group (75.7% v. 68.0%, OR = 1.48, 95% CI 1.12–1.96, P = 0.006). Even an increase in the minimal improvement to 50% did not change the result of a significantly more positive effect of heroin treatment (69.5% v. 58.6%, OR = 1.63, 95% CI 1.26–2.13, P < 0.001). The analysis with a single response criterion – those participants responding on both primary outcome measures – allows for an easier comparison with and confirmation of the Dutch results. However, the analysis of separate response criteria has the advantage of allowing a more differentiated analysis of effects.

Another aspect that needs to be discussed is the improvement in the month between baseline and initiation of treatment with study medication, especially with respect to physical health. This improvement is probably the result of a combination of regression to the mean and treatment between baseline assessment and randomisation. Considering the very poor health status of the sample at baseline, for ethical reasons physical and/or mental health problems had to be attended to even before initiation of study treatment. However, since randomisation took place thereafter, treatment prior to randomisation and possible improvements do not bias the observed differences between the two medication conditions (heroin or methadone) after 12 months’ treatment. None the less, if the response criteria for physical health were defined using the OTI score at initiation of treatment as the baseline, 77.1% of the heroin group and 69.2% of the methadone group would have been defined as responders for the primary outcome measure ‘health’, with a significant difference (OR = 1.50, 95% CI 1.13–1.99, P = 0.005).

The rather low retention rate in the methadone group could be considered a further limitation. The high drop-out rate in the methadone group is probably a result of the disappointment at not being randomised into the heroin group. However, a large portion of those dropping-out took up other treatments, so that the limiting effect of the low retention rate is minimised in a randomised intent-to-treat analysis.

A final limitation is that not all data on illicit drug use were based on objective urine or hair analysis, self-reported data were also included. However, studies have shown self-reported data to be accurate, reliable and valid, provided that confidentiality is ensured and no sanctions are connected to the answers (Rounsaville, 1993).

Implications

This large multicentre study confirms the results of the Swiss (Rehm et al., 2001) and Dutch (van der Brink et al., 2003) studies and therefore addresses the limitations pointed out by the Cochrane review (Ferri et al., 2005) by providing strong further evidence of the efficacy of prescribed heroin in the treatment of people with opioid dependence who have not profited from other forms of treatment. Considering the higher rate of serious adverse events, heroin prescription should remain a treatment of last resort for people who are currently or have in the past failed at maintenance treatment.

ACKNOWLEDGEMENTS

We thank patients and staff who participated in the study, as well as the Safety and Advisory Boards for their advice and support. The trial was commissioned and funded by a joint working group of the German Ministry of Health, the seven participating cities and the states of Hessen, North Rhine-Westphalia and Lower Saxony.

REFERENCES


Brain opioid receptor binding in early abstinence from opioid dependence

Positron emission tomography study

TIM M. WILLIAMS, MARK R. C. DAGLISH, ANNE LINGFORD-HUGHES, LINDSAY G. TAYLOR, ALEXANDER HAMMERS, DAVID J. BROOKS, PAUL GRASBY, JUDITH S. MYLES and DAVID J. NUTT

Background  Although opioid receptor function in humans is clearly reduced during opioid dependence, what happens to the receptor in early abstinence is not understood.

Aims This study sought to examine changes in opioid receptor availability in early abstinence from opioid dependence.

Method Ten people with opioid dependence who had completed in-patient detoxification and 20 healthy controls underwent [11C]-diprenorphine positron emission tomography (PET) to measure levels of available opioid receptors in the brain of patients on methadone maintenance, but found no detectable occupancy by methadone (Melichar et al., 2005). Other PET studies have reported that increased binding of [11C]-carfentanil in withdrawal and abstinence in cocaine and alcohol dependence is associated with craving (Zubieta et al., 1996; Gorelick et al., 2005; Heinz et al., 2005). A preliminary study also reported an increase in [11C]-carfentanil binding in people with opioid dependence who were briefly maintained on buprenorphine (Zubieta et al., 2000). In this study we present data on the binding of the opioid receptor PET tracer [11C]-diprenorphine, which labels μ, κ and δ opioid receptors, in people with opioid dependence during early abstinence. We measured [11C]-diprenorphine binding in brain areas implicated in dependence and its relationship to clinical variables. Our hypothesis was that in people with opioid addiction, opioid receptor availability would be increased in early abstinence and that this would be related to craving.

Results Compared with controls, participants with opioid dependence had increased [11C]-diprenorphine binding in the whole brain and in 15 of the 21 a priori regions studied.

Conclusions This study suggests that opioid receptor binding is increased throughout the brain in early abstinence from dependent opioid use. These data complement the findings in cocaine and alcohol dependence.

Declaration of interest None. Funding detailed in Acknowledgements.
Heroin Craving Questionnaire (HCQ; Feinstein et al., 1997) and the Obsessive Compulsive Drinking Scale (OCDS), adapted to measure opioid compulsive behaviour and obsessive thoughts, and allowing for mode of drug delivery (Anton et al., 1996). Two experienced addiction clinicians (T.W. and M.D.) independently rated each patient for the amount of opioids used in the month and year prior to scanning, as well as lifetime use, using a structured rating scheme taking into account the combination of opioids used, length and route of use. We also looked for associations with published post-mortem data reporting regional densities of μ opioid receptors (Pfeiffer et al., 1982).

Participants also completed the Spielberger State–Trait Anxiety Inventory (STAI; Spielberger, 1983), the 36-item short form of the General Health Survey (SF-36; McHorney et al., 1994), the revised Eysenck Personality Questionnaire (EPQ–R; Eysenck & Eysenck, 1975) and the Eysenck Impulsiveness Questionnaire (IVE; Eysenck et al., 1985).

**PET scans**

All participants underwent [11C]-diprenorphine PET using a CTI/Siemens (Munich, Germany) ECAT 953b brain camera with high-sensitivity three-dimensional mode. A bolus of 370 MBq [11C]-diprenorphine was given intravenously over 30 s. Dynamic emission data were acquired over 90 min, in 18 time frames (27 frames for 9 controls) and reconstructed into 31 contiguous horizontal image planes (Jones et al., 1994). Radioactivity in arterial blood was assayed continuously online in accordance with a standard protocol and discrete blood samples were taken every 5–10 min for assay of radiolabelled metabolites in plasma (Ranicar et al., 1991).

**Image processing and statistical analysis**

The dynamic PET scans were analysed to produce parametric images of ligand volume of distribution using spectral analysis with in-house receptor parametric mapping software implemented in Matlab (Mathworks Inc., Natick, Massachusetts) (Gunn et al., 2002). Spectral analysis with individual metabolite corrected plasma input function takes account of any difference in tracer delivery between individuals or groups. The volume of distribution is the ratio of total free and bound tissue to free plasma ligand concentration at equilibrium and provides an index of receptor binding. We used Statistical Parametric Mapping (SPM2, Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, UK) running in Matlab to transform the volume of distribution images into a standard space as defined by the Montreal Neurological Institute (Evans et al., 1993) using a weighted-mean [11C]-diprenorphine template created inhouse from the PET scans of seven healthy volunteers.

We performed two types of analysis, first using predefined volumes of interest and then using SPM2 for an exploratory analysis. Twenty-one regions for which we had an a priori hypothesis for increased opioid receptor availability based on previous studies were selected for comparison. These areas were the orbito-frontal cortex, anterior cingulate, ventral striatum (including the nucleus accumbens), dorsal striatum (including the caudate nucleus and putamen), thalamus, amygdala and periaqueductal grey matter. All these regions have been shown to have a role in the addiction process. The anterior cingulate cortex, orbitofrontal cortex, nucleus accumbens and amygdala are key in reward and motivation during drug-use from the evaluation of stimuli to reward-based decision-making and learning. The periaqueductal grey matter is an important element of the endogenous opioid system and is involved in conditioned processes in dependent drug use. Similarly the thalamus, caudate and putamen form part of the emotional reward neurocircuitry which has an important role in motivational factors and links to motor pathways, possibly being a route for the development of locomotor sensitisation with continued drug use (Kalivas & Volkow, 2006; Nutt et al., 2006).

We used statistical parametric mapping to transfer 73 standardised volumes of interest derived from a probabilistic atlas of brain images (Hammers et al., 2003) onto individual scans by inverting the deformations used to spatially normalise the images. The volume of distribution maps were sampled using the individualised atlas for every participant to generate mean volume of distribution values for each volume of interest. These values were then compared between groups using a twotailed non-paired t-test – unequal variances were assumed. Pearson’s correlation statistics were used to assess the association of clinical variables with opioid receptor binding.

In addition, [11C]-diprenorphine volume of distribution images were analysed on a voxel-by-voxel basis using SPM2. Spatially normalised parametric images were smoothed with a 12 mm kernel at full width half maximum. Mean differences between groups were interrogated using non-paired t-tests, and correlations between clinical variables and [11C]-diprenorphine binding were explored using linear regression within the general linear model in SPM2. Proportional scaling was used to normalise global differences throughout. For regions where there existed an a priori hypothesis, results are reported as significant at a threshold of $p < 0.05$ uncorrected. For all other areas familywise error correction was used.

**Table 1** Drug and alcohol use in the 30 days prior to scanning

<table>
<thead>
<tr>
<th>Drug</th>
<th>Ever used (n=10)</th>
<th>Any use in past 30 days (n=10)</th>
<th>Days used in past 30 days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (s.d.)</td>
<td>Median</td>
<td>Range</td>
</tr>
<tr>
<td>Heroin</td>
<td>10</td>
<td>14.7 (4.7)</td>
<td>17</td>
</tr>
<tr>
<td>Intravenous</td>
<td>10</td>
<td>12.7 (6.4)</td>
<td>16</td>
</tr>
<tr>
<td>Smoked</td>
<td>10</td>
<td>17.5 (3.9)</td>
<td>17.5</td>
</tr>
<tr>
<td>Cannabis</td>
<td>10</td>
<td>10.3 (7.7)</td>
<td>10.5</td>
</tr>
<tr>
<td>Crack</td>
<td>9</td>
<td>8.2 (5.4)</td>
<td>6</td>
</tr>
<tr>
<td>Benzodiazepines</td>
<td>10</td>
<td>15.0 (6.0)</td>
<td>16</td>
</tr>
<tr>
<td>Amphetamines</td>
<td>9</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>MDMA</td>
<td>8</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>LSD</td>
<td>10</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>Alcohol</td>
<td>10</td>
<td>8.2 (6.3)</td>
<td>5</td>
</tr>
<tr>
<td>Nicotine</td>
<td>10</td>
<td>0</td>
<td>30</td>
</tr>
</tbody>
</table>

MDMA, methylenedioxyamphetamine; LSD, lysergic acid diethylamide.
RESULTS

Clinical measures

At the time of scanning, there was no clinically significant opioid withdrawal as measured by the OWS (mean score 1.2, s.d.=1.2, range 0–3). However the subjective ARCI and Adjective Checklist showed higher measures of withdrawal in participants with opioid dependence than in controls. They showed significantly higher state, but not trait, anxiety than controls on the day of their scan. Participants with opioid dependence scored significantly lower than controls on some measures of health (Table 2). On the personality questionnaire participants with opioid dependence scored significantly higher for psychotism, extraversion, impulsiveness and venturesomeness than controls, but not for neuroticism or empathy.

All participants with opioid dependence reported craving on the HCQ (mean score 15.7, s.d.=6.0) and modified OCDS (mean score 21.67, s.d.=10.6), and the scores were highly correlated (r=0.76, P<0.018). Craving scores were comparable with our previous study of the same stage of detoxification where craving was elicited using an imagery-based procedure (Weinstein et al, 1997).

Image analysis

Participants with opioid dependence showed a significantly higher level of opioid receptor availability, as measured by global [11C]diprenorphine volume of distribution, when compared with controls (19.3 ± 17.1, 11.4% increase, 95% CI 2.1–20.7, P=0.019). In 15 of the 21 a priori regions studied there was a significant increase in volume of distribution in people with opioid dependence when compared with controls (P<0.05 uncorrected). These were the brain-stem, right amygdala, left medial orbital cortex and bilateral anterior cingulate, putamen, thalamus, and anterior, lateral and posterior orbitofrontal cortex. Only the left lateral orbital area remained significant if these areas are considered independent and the overly conservative Bonferroni correction is applied (P=0.042, corrected). There was no significant association between age and global [11C]diprenorphine volume of distribution for the whole group (n=30, r=−0.30, P=0.105) or when the control (n=20, r=−0.31, P=0.184) and opioid-dependent groups (n=10, r=−0.02, P=0.962) were analysed separately.

Table 2  Clinical measures

<table>
<thead>
<tr>
<th>Test and measure</th>
<th>Opioid dependent (n=10)</th>
<th>Controls (n=8)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Withdrawal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer Withdrawal Scale</td>
<td>1.2 (1.2)</td>
<td>0.1 (0.4)</td>
<td>0.024</td>
</tr>
<tr>
<td>Adjective Checklist–withdrawal</td>
<td>20.8 (9.9)</td>
<td>5.0 (5.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Adjective Checklist – agonist effects</td>
<td>22.5 (9.4)</td>
<td>20.5 (5.1)</td>
<td>0.595</td>
</tr>
<tr>
<td>ARCI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Withdrawal effects</td>
<td>23.0 (9.9)</td>
<td>5.0 (5.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Euphoric effects</td>
<td>7.8 (3.4)</td>
<td>4.3 (2.8)</td>
<td>0.030</td>
</tr>
<tr>
<td>Agonist effects</td>
<td>22.5 (9.3)</td>
<td>20.5 (5.1)</td>
<td>0.595</td>
</tr>
<tr>
<td>Spielberger</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSAI–state</td>
<td>43.7 (8.6)</td>
<td>32.9 (9.1)</td>
<td>0.014</td>
</tr>
<tr>
<td>STAI–trait</td>
<td>44.5 (11.5)</td>
<td>37.5 (10.6)</td>
<td>0.172</td>
</tr>
<tr>
<td>SF–36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General health</td>
<td>43.6 (19.5)</td>
<td>82.5 (14.2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Physical function</td>
<td>80.2 (10.6)</td>
<td>96.9 (5.3)</td>
<td>0.001</td>
</tr>
<tr>
<td>Physical role limitation</td>
<td>45.0 (36.9)</td>
<td>96.9 (8.8)</td>
<td>0.001</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>39.7 (17.9)</td>
<td>83.4 (19.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Emotional role limitation</td>
<td>43.3 (41.7)</td>
<td>87.5 (35.4)</td>
<td>0.030</td>
</tr>
<tr>
<td>Social functioning</td>
<td>52.5 (23.4)</td>
<td>87.6 (17.7)</td>
<td>0.003</td>
</tr>
<tr>
<td>Vitality</td>
<td>39.0 (16.5)</td>
<td>49.6 (16.5)</td>
<td>0.193</td>
</tr>
<tr>
<td>Mental health</td>
<td>43.8 (11.8)</td>
<td>45.8 (9.9)</td>
<td>0.699</td>
</tr>
<tr>
<td>EPQ–R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychotism</td>
<td>11.0 (2.3)</td>
<td>5.2 (2.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Extraversion/intraversion</td>
<td>18.9 (4.4)</td>
<td>13.7 (3.0)</td>
<td>0.008</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>15.3 (5.1)</td>
<td>11.9 (8.2)</td>
<td>0.285</td>
</tr>
<tr>
<td>IVE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impulsiveness</td>
<td>13.5 (4.5)</td>
<td>4.9 (2.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Venturesomeness</td>
<td>12.3 (2.8)</td>
<td>8.0 (3.5)</td>
<td>0.010</td>
</tr>
<tr>
<td>Empathy</td>
<td>13.0 (4.0)</td>
<td>14.3 (3.1)</td>
<td>0.477</td>
</tr>
</tbody>
</table>

1. n=10.
2. n=9.

ARCI, Addiction Research Centre Inventory; SF–36, 36-item short form of the General Health Survey; EPQ–R, Eysenck Personality Questionnaire – Revised; IVE, Eysenck Impulsiveness Questionnaire.

Fig. 1  Global and a priori regional [11C]-diprenorphine binding for people with opioid dependence (■) and controls (■). *P<0.05 unpaired t-test.
The re-detoxification daily methadone total dose (mg/kg) and duration of methadone use showed no association with either global or regional $[^{11}C]$-diprenorphine binding. Furthermore, no relationship with $[^{11}C]$-diprenorphine binding was found when the participants with opioid dependence were divided into two groups: short-term users (use for under 10 years) and long-term users (use for more than 10 years). We found no significant effect for either alcohol or ‘crack’ cocaine on $[^{11}C]$-diprenorphine binding by comparing those that had used in the previous 30 days with those that had not. Of interest, although it did not achieve significance, was a trend towards a negative correlation between level of opioid use in the previous month, year or lifetime and global $[^{11}C]$-diprenorphine binding. There was no correlation between $[^{11}C]$-diprenorphine volume of distribution and craving, subjective opioid effects or withdrawal, or any of the personality variables.

Regional densities of $\mu$ opioid receptors, as reported in post-mortem tissue (Pfeiffer et al., 1982), correlated with $[^{11}C]$-diprenorphine binding in each region for the whole group ($r=0.54$, $P=0.006$), and when control and patient groups were analysed separately ($r=0.51$, $P=0.010$ and $r=0.53$, $P=0.007$ respectively). There was no correlation between $[^{11}C]$-diprenorphine volume of distribution and $\kappa$ opioid receptor (combined group $r=0.05$, $P=0.803$) or $\delta$ opioid receptor regional densities (combined group $r=0.31$, $P=0.141$). No differences were apparent between the groups in these correlations.

An exploratory comparison of people with opioid dependence and controls using statistical parametric mapping showed a significant increase in $[^{11}C]$-diprenorphine binding only in the right fusiform and parahippocampal gyri (MNI coordinates $x=38$ mm, $y=−26$ mm, $z=−32$ mm, cluster size 594 voxels, peak $T=5.15$, $P=0.016$ familywise error corrected). No significant correlations were found using statistical parametric mapping between $[^{11}C]$-diprenorphine volume of distribution and any clinical variables. Using a statistical threshold of $P<0.05$ uncorrected to investigate the a priori areas confirmed the findings of the region of interest analysis.

**DISCUSSION**

This is the first study, to our knowledge, that has studied opioid receptor binding using $[^{11}C]$-diprenorphine PET in people with opioid dependence undergoing recent detoxification. We found a significant increase in $[^{11}C]$-diprenorphine volume of distribution in the majority of a priori regions of interest, although notably not in the nucleus accumbens or caudate nucleus. Exploratory statistical parametric mapping at a threshold for significance applying a full correction for multiple comparisons identified the right fusiform and right parahippocampal gyri as areas of significantly increased binding. At a lower threshold, the mapping confirmed the findings of the volume of interest analysis when applied to regions specified a priori.

**Opioid receptor availability in dependence**

There is limited preclinical research that helps to interpret the results of this study. Although it is clear that chronic opioid exposure leads to reduced opioid receptor function (tolerance), the mechanisms through which this is achieved are not certain and may include receptor internalisation or reduced receptor–effector coupling (Williams et al., 2001). In vitro studies have shown that chronic exposure to an opioid agonist can lead to a downregulation in opioid receptors (Goodman et al., 1996). However, this is not a consistent pattern in in vivo studies, which have reported increases, decreases and no change in opioid receptors, depending on the paradigm used (Zadina et al., 1995). In humans, tolerance to opioid agonists is well characterised but there are virtually no data on brain opioid receptor imaging. We have previously demonstrated a dose-related reduction in opioid receptor function in people with opioid dependence who are on methadone maintenance by showing that they are less sensitive to the effects of an opioid agonist, hydromorphone (Melchior et al., 2003). However, in a parallel $[^{11}C]$-diprenorphine PET study, we found no difference in binding compared with a healthy control group, suggesting limited occupancy and no significant changes in receptor number (Melchior et al., 2005). This complements a study using $[^{18}F]$-cycloxylo PET that also suggested that methadone requires only very low levels of opioid receptor occupancy for efficacy (Kling et al., 2000). Lastly, post-mortem studies of people with heroin dependence have shown inconsistent changes (reduction or no difference) in $\mu$ opioid receptor density compared with healthy controls (Gabilondo et al., 1994; Ferrer-Alcon et al., 2004). These studies suggest that chronic opioid exposure might not alter opioid receptor availability and importantly not increase receptor availability.

**Increased $[^{11}C]$-diprenorphine binding**

The increase in $[^{11}C]$-diprenorphine binding reflects an increase in availability of opioid receptors to this PET tracer. Increased receptor affinity for the tracer could account for this increased availability, but there is no preclinical evidence that chronic opioid administration alters affinity. Therefore, the increase in $[^{11}C]$-diprenorphine binding might be due to: (a) an increase in opioid receptors during early abstinence from opioid drugs; (b) an increase in opioid receptor number that develops with the chronic use of an opioid agonist; (c) a reduction in competition from endogenous opioids. We believe that it is most likely that our findings reflect a significant increase in opioid receptor number immediately following detoxification from opioids. We know that withdrawal and early abstinence is a time when the brain is under stress, and that such an increase might represent a neuroadaptive response. This would explain the similar findings after cocaine and alcohol dependence (Zubieta et al., 1996; Gorelick et al., 2005; Heinz et al., 2005).

Increased $[^{11}C]$-diprenorphine binding could also reflect increased opioid receptor availability as a result of suppression of endogenous opioid release. Preclinical evidence shows that chronic treatment with methadone does not alter the concentration or function of endogenous opioids, although later studies with other opioids and other drugs of misuse suggest that endogenous opioids play a role in craving or drug-seeking behaviour (for a review see Gerrits et al., 2003). Activation of the endogenous opioid system is associated with the regulation of emotions, physical and emotional pain (Ribeiro et al., 2005). A possibility is that the exogenous opioids used to alter emotions by people with opioid dependence might lead to suppression of the endogenous opioid system and consequently a compensatory upregulation of receptors. This would leave more receptors available for occupancy by $[^{11}C]$-diprenorphine in early abstinence. We are not aware of any human studies describing the impact of chronic opioid
agonist use on levels of endogenous opioids.

**Opioid system after abstinence from substances**

In addition to being the primary target for opioid drugs, the opioid neurotransmitter system is important in initiating and maintaining dependence on a variety of misused substances (Herz, 1997; Gerrits et al., 2003; Kreek et al., 2004). A number of recent neuroimaging studies in humans using the \( \mu \)-selective agonist \(^{11}C\)-carfentanil have reported elevations of tracer binding in early abstinence from cocaine and alcohol, which are associated with craving (Zubieta et al., 1996; Gorelick et al., 2005). Detoxification from a short course of buprenorphine has been shown in a preliminary study to result in a significant increase in \( \mu \) opioid \(^{11}C\)-carfentanil binding in the inferofrontal cortex and anterior cingulate regions compared with controls (Zubieta et al., 2000). Therefore, it appears that similar increases in opioid receptor availability are seen during early abstinence from cocaine and alcohol, and preliminary data suggest a comparable increase in people with opioid dependence.

The evidence to date points to elevations in opioid binding being an acute effect of early abstinence, and our results in opioid dependence complement these findings. It is not clear whether these changes persist or even become additive with progressive detoxifications. In cocaine dependence, opioid receptor binding in some but not all regions returns to control levels within 1 week (Gorelick et al., 2005). In alcohol dependence the increase appears more persistent, with no reduction evident after 5 weeks of abstinence (Heinz et al., 2005). It was not possible to scan our participants after a period of abstinence owing to high relapse rates and strict residential rehabilitation programmes, but this would be valuable in future studies.

We found significant increases in \(^{11}C\)-diprenorphine binding in the majority of regions analysed using the atlas, and significant increases in fusiform/parahippocampal gyri using exploratory voxel-based statistical parametric mapping, although increases were seen throughout the brain when the threshold for significance was lowered. It is not clear why an area incorporating the fusiform/parahippocampal gyri which is involved in processing visual associations and memory was highlighted by statistical parametric mapping. We found no difference in \(^{11}C\)-diprenorphine binding between people with opioid dependence and controls in several of the \textit{a priori} areas, notably the periaqueductal grey matter (in the brain-stem), nucleus accumbens and caudate. The template used for the brain-stem region is not precise enough to isolate the periaqueductal grey matter within the brain-stem region of interest, which may account for the lack of association there. However, we are surprised to find no association with the nucleus accumbens and caudate in the light of previous findings of increased receptor number in these areas during withdrawal from cocaine and alcohol. In the two studies of cocaine dependence, significant increases were seen in the ventral striatum and the anterior cingulate, frontal and temporal cortices, caudate and thalamus (Zubieta et al., 1996; Gorelick et al., 2005), whereas in alcohol dependence, significant increases were restricted to the ventral striatum (Heinz et al., 2005). In people with opioid dependence the changes were much more widespread, perhaps because of the direct pharmacological effect of opioids and possible changes in the endogenous opioid system.

**Opioid receptor availability and clinical variables**

We found no correlation between craving and opioid receptor availability, which is at variance with our hypothesis and previous findings in alcohol and cocaine dependence (Zubieta et al., 1996; Gorelick et al., 2005; Heinz et al., 2005). Our participants with opioid dependence demonstrated high scores on two rating scales for craving, which were comparable with those in a previous study (Weinsteind, 1997) and with individuals maintained on methadone who had withdrawal induced by naloxone (Schuster et al., 1995) but were higher than scores for people maintained on methadone (Schuster et al., 1995). Furthermore, our participants experienced levels of pain and variance in craving scores that were comparable with earlier studies in which craving was induced and resulting brain activation detected (Daglish et al., 2001). Craving measures vary and so comparison with other studies is hampered. However, in our study we chose two commonly used scales with a total of seven craving subscales, so the absence of a correlation here is robust. In other studies reporting a relationship between craving and opioid receptor levels, \(^{11}C\)-carfentanil, a \( \mu \)-selective tracer was used (Zubieta et al., 1996, 2000; Gorelick et al., 2005; Heinz et al., 2005). It may be that since \(^{11}C\)-diprenorphine labels \( \mu, \kappa \) and \( \delta \) opioid receptors, \( \mu \) receptor-related changes were obscured by alterations in the other subtypes. However we think this unlikely as the \(^{11}C\)-diprenorphine signal correlated only with the reported \( \mu \) opioid receptor density in each brain region and not with the \( \kappa \) and \( \delta \) opioid receptor density. Nevertheless, it would be beneficial to repeat this study using a more selective opioid receptor tracer, such as \(^{11}C\)-carfentanil, to determine whether the increase in opioid receptor binding demonstrated here is mainly due to increase in any particular subtype.

Opioid receptor binding levels were not related to withdrawal symptoms as found in cocaine and alcohol dependence (Zubieta et al., 1996; Gorelick et al., 2005). This is consistent with the clinical picture where opioid withdrawal can be ameliorated by non-opioid pharmacotherapy. We did not find a correlation between age and opioid receptor levels in either the group with opioid dependence or controls. In a \(^{11}C\)-carfentanil PET study of healthy controls with a wider age range, increasing age was associated with higher opioid receptor levels in the neocortex (Zubieta et al., 1999). Our more limited age range and younger average age likely contributed to the absence of such a correlation. All of our group with dependence were tobacco smokers and controls were current non-smokers, but there was no correlation between quantity of cigarettes smoked and \(^{11}C\)-diprenorphine binding. Furthermore, another study of alcoholism reported no significant interaction between smoking status and \( \mu \) opioid receptor availability in patients and controls (Heinz et al., 2005).

**Limitations**

Although this study was appropriately powered to detect the measured effect in a PET study of this nature, it may have been underpowered to determine associations with clinical variables, especially craving. The studies reporting an association between craving and opioid receptor levels had dependent groups of 10, 17 and 25 respectively (Zubieta et al., 1996; Gorelick et al., 2005; Heinz et al., 2005). However, the participants in our study were craving at

67
similar levels and with a wide range of craving scores, making it likely that any association should have been apparent.

Implications
We have reported a significant widespread increase in brain opioid receptor availability in people with opioid dependence during early abstinence from methadone. Together with previous evidence, we argue that this reflects an increase after cessation of methadone rather than a chronic change. If this is the case, it could give us a crucial insight into the mechanisms that underlie opioid craving. Although clinically we know that substitution treatment is effective we do not know whether prolonged agonist exposure permanently alters brain neurochemistry and whether these changes hamper recovery. Furthermore, since such an increase in opioid receptors has also been shown in alcohol and cocaine dependence, this argues for a fundamental role of the opioid system in addiction, or at least in the early abstinence syndrome. The contribution of this to clinical states and treatment outcomes has yet to be fully characterised.

ACKNOWLEDGEMENTS
This work was funded by a programme grant from the Medical Research Council, UK, and by Avon and Wiltshire Partnership NHS Trust research and development funding. Radiotracers and scanners were supplied by Hammesmacher Imnat, GE Healthcare.

REFERENCES
regulation through the endogenous opioid system and μ-opioid receptors. Progress in Neuropsychopharmacology and Biological Psychiatry, 29, 1264–1280.


Long-term outcomes after discharge from medium secure care: a cause for concern

STEFFAN DAVIES, MARTIN CLARKE, CLIVE HOLLIN and CONOR DUGGAN

Background There are few long-term follow-up studies of patients discharged from medium secure units in the UK, even though these units were introduced over 20 years ago.

Aims To describe mortality, rates of reconviction at different time periods; violent behaviour (not leading to conviction), readmission and employment, after discharge from a medium secure unit.

Method Of 595 first admissions over a 20-year period, 550 discharged cases were followed-up. Multiple data sources were used.

Results Fifty-seven (10%) patients had died, of whom 18 (32%) died by suicide, and the risk of death was six times greater than in the general population. Almost half (49%) of those discharged were reconvicted and almost two-fifths (38%) of patients were readmitted to secure care.

Conclusions Community psychiatric services need to be aware that those discharged from medium secure care are a highly vulnerable group requiring careful follow-up if excess mortality, high levels of psychiatric morbidity and further offending are to be prevented.

Declaration of interest None. Funding detailed in Acknowledgements.

Although the psychiatric management of specific individuals has often had a disproportionate effect on mental health policy in the UK, data on the outcome of the generality of patients from secure psychiatric services are surprisingly limited, although there are low-volume and high-cost forms of care. Medium secure services are a case in point, with few data whereby their efficacy can be judged. In one of the few published studies, Maden et al (1999) found that in 24% of 234 first admissions the patient had been reconvicted at a mean of 6.6 years, 9% of patients had died and 74% had been readmitted. Reconviction in a larger sample (n=959) followed up for 2 years was 15% (Maden et al, 2004). Given the limited nature of previous work, we sought to extend it in this study by examining the fate of an entire first admission cohort to medium security across a range of outcomes and services, over a 20-year period, using multiple sources of information.

METHOD

Sample The study sample comprised all admissions to the Arnold Lodge medium secure unit in Leicester since its opening in July 1983 up to 30 June 2003 – a period of 20 years. Ricketts et al (2001) previously reported on the characteristics of the 504 patients admitted to this unit up to 30 June 1999. Those data were checked again as part of this study and a further 8 patients were identified, together with an additional 83 patients who were admitted between 1 July 1999 and 30 June 2003. Thus, there were in total 595 first admissions to Arnold Lodge over the 20 years since the unit opened.

Data sources Admission characteristics were derived from medical records at Arnold Lodge. Outcome data were obtained from clinical records at Arnold Lodge, other psychiatric services, the Home Lodge Mental Health Unit, the Office for National Statistics (ONS), the general practitioner registra-tions database, the Offenders’ Index and the Police National Computer (for reconvictions), the electoral roll (UK-Info Disk version 10, i-CD Publishing, London, UK) and the LexisNexis database of newspaper reports.

Outcome measures A proforma was designed to record several outcomes, including data on reconviction, psychiatric contact, accommodation and psychosocial variables for each year of follow-up for each case in the study. In this paper we report on the sample’s mortality, reconviction, behaviours not resulting in conviction, readmission to secure and open hospitals, employment and accommodation.

Mortality Death certificates were obtained from the ONS. The mortality of the sample overall was compared with that of the general population by indirect standardisation to the England and Wales mortality rates for 10-year age bands published by the ONS (http://www.statistics.gov.uk). This yielded a standardised mortality ratio together with its 95% confidence interval.

Classifying reconvictions Although over 30 forensic follow-up studies have been published to date, none has used the Home Office standard method for reporting reconviction. This classifies convictions as either ‘grave’ or ‘standard list’ offences (Home Office, 2002). Grave offences are those for which the maximum sentence is life imprisonment, plus arson not endangering life, and include murder, attempted murder, robbery, rape and arson. Standard list offences are all other indictable offences tried in either a Crown court or a magistrates' court. The Home Office also standardises its reporting of reconviction, for example at 2 years and 5 years following the date of release from prison. The authors have adopted this classification because it allows comparison with criminal justice statistics.

In line with Home Office practice, the date of the conviction rather than the actual date of the offence was used in the time to reconviction analyses because the date of the actual offence was not available for all
offences, although it is accepted that this is a conservative approach that is likely to underestimate the rate of reoffending. The Police National Computer database provides the date when a person was charged, but this information was not available in all cases. The time to reconviction presented in the study was calculated from the point of discharge from Arnold Lodge rather than time of entering the community. However, where the case notes or reports were available, violent and aggressive episodes and fire-setting were recorded including those by patients in hospitals, prison or the community.

**Ethical considerations**

Ethical approval was granted from the Trent Multicentre Research Ethics Committee. In the light of the known difficulty in both identifying and gaining the consent of forensic psychiatric patients, the research was conducted under section 60 of the Health and Social Care Act 2001. This permits the use of identifiable National Health Service (NHS) patient information under certain circumstances, without the consent of patients. This was the first study of a psychiatric population to be granted section 60 approval. Statistical analyses were conducted using the Statistical Package for the Social Sciences, version 11.5 for Windows.

**RESULTS**

The catchment area originally served by the Arnold Lodge unit was the former Trent Region comprising the counties of Leicestershire, Nottinghamshire, Derbyshire, Lincolnshire and South Yorkshire (population approximately 4.8 million according to ONS estimates for mid-2001). However, this area was reduced in 1997 – 14 years into the follow-up – to Leicestershire, Nottinghamshire, South Derbyshire and Lincolnshire (approximate population 3.3 million) when another medium secure unit opened in the north of the region.

Of the 595 first admissions, there were 502 men (84.4%) and 93 women (15.6%). Four people had died during their admission, 550 people had been discharged and 41 people had yet to be discharged at the census date. Hence, 554 ‘discharges’ constituted the sample used in the analyses (apart from mortality, which included all admissions). The mean length of stay for this sample was 346 days (s.d.=468.2), ranging from 2 days to 3872 days. The mean age on admission was 29.9 years (s.d.=9.1).

The Mental Health Act 1983 classification of these admissions comprised 67.2% mental illness, 26.6% psychopathic disorder, 3.0% mental illness and psychopathic disorder, and 0.5% mental impairment; the classification for 2.4% was unknown and 0.3% did not have a classification.

The mean length of follow-up from discharge to death, loss of contact, or the census date was 9.4 years (s.d.=4.8). Women had a longer mean follow-up time than men – 11.5 years (s.d.=4.1) and 9.0 years (s.d.=4.8) respectively (t=5.062, d.f.=144.9, P<0.001). There was no significant difference in the mean follow-up times between patients with a Mental Health Act classification of either psychopathic disorder (9.8 years, s.d.=5.1) or mental illness (9.2 years, s.d.=4.6). Overall there were 5771 person-years of follow-up from admission (including mortality in the unit) and 5264 person-years from discharge.

**Discharge location**

Of the 534 ‘discharges’, 34.3% were discharged to a psychiatric hospital of a lower security, predominantly open wards; 27.3% were discharged to the community (which includes home or a hostel); 26.5% of patients were transferred to the criminal justice system, either returned to prison or to court for sentencing; 7.2% were transferred to high secure care; 2.9% were transferred to a different medium secure unit; 0.7% died while in the unit; and the discharge location was unknown for 1.1%, mainly due to their being discharged in their absence after going absent without leave or failing to return from leave.

**Mortality**

At the census, whether the individual was alive or dead was known for 522 of the 550 discharged patients (95%). However, 23 of the remaining 28 patients were confirmed as being alive to at least 2000 (data from electoral rolls, general practitioner registrations and the Police National Computer). Four patients died during their first admission and a further 53 patients died after discharge. The crude risk of death was 9.6% overall (9.2% for men and 14.6% for women) with a mean age at death of 43.6 years (s.d.=12.9). Only 25 deaths (44%) had a verdict of natural causes. There were 18 deaths from suicide (32%) and 14 (25%) from other unnatural causes (Table 1).

Standardised mortality ratios (SMRs) were calculated from admission rather than discharge so as to include the four individuals who died in hospital during their admission. The risks of death for men, women, the whole cohort and deaths by different causes were all significantly higher than those expected in the general population (Table 2). For instance, the risk of death was 6 times greater than expected for the whole cohort, almost 19 times greater for deaths from unnatural causes and over 32 times greater for deaths from suicide. The SMRs for the Mental Health Act legal classifications of mental illness and psychopathic disorder were 6.3 and 4.6 respectively.

**Reconviction**

Almost half (48.7%) of those discharged were reconvicted of an offence over the entire period of follow-up (264 of 542) reconviction data for 8 patients were available and 2 were excluded. In total, 2 of the 8 patients who reoffended were women.

**Table 1 Cause of death**

<table>
<thead>
<tr>
<th>Cause of death</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural</td>
<td>71</td>
</tr>
<tr>
<td>Cancer</td>
<td>5</td>
</tr>
<tr>
<td>Bronchopneumonia</td>
<td>6</td>
</tr>
<tr>
<td>Heart-related</td>
<td>9</td>
</tr>
<tr>
<td>Other (e.g. obesity, peritonitis)</td>
<td>5</td>
</tr>
<tr>
<td>Suicide</td>
<td></td>
</tr>
<tr>
<td>Hanging</td>
<td>10</td>
</tr>
<tr>
<td>Drowning</td>
<td>3</td>
</tr>
<tr>
<td>Overdose/poisoning</td>
<td>3</td>
</tr>
<tr>
<td>Other (e.g. fall from height)</td>
<td>2</td>
</tr>
<tr>
<td>Open verdict</td>
<td></td>
</tr>
<tr>
<td>Hanging</td>
<td>3</td>
</tr>
<tr>
<td>Collusion with train</td>
<td>1</td>
</tr>
<tr>
<td>Fall from building</td>
<td>1</td>
</tr>
<tr>
<td>Poisoning</td>
<td>1</td>
</tr>
<tr>
<td>Collapse after medication</td>
<td>1</td>
</tr>
<tr>
<td>Organ failure/peritonitis</td>
<td>1</td>
</tr>
<tr>
<td>Misadventure</td>
<td></td>
</tr>
<tr>
<td>Overdose/poisoning</td>
<td>2</td>
</tr>
<tr>
<td>Swallowed foreign body</td>
<td>1</td>
</tr>
<tr>
<td>Accidental</td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>2</td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
</tr>
<tr>
<td>Head injuries</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
</tr>
</tbody>
</table>
missing). The locations where the offences were committed were as follows: community 225 (83%), hospital 24 (9%) and prison 5 (2%); the location was not known for 10 (4%). The mean time from discharge to first conviction was 3.2 years (s.d.=3.2, median=1.9); the maximum time to conviction after discharge was 16.4 years.

Table 3 shows the number of patients who were convicted of a standard list or grave offence at 2 years, at 5 years and at any point during the follow-up. Patients without a conviction but who did not have a full 2-year or 5-year follow-up were excluded. The only significant difference for reconviction according to gender showed that men were more likely than women to have been convicted of a standard list offence in the first 5 years after discharge ($\chi^2=7.0$, d.f.=1, $P=0.008$). The only significant difference for reconviction by Mental Health Act classification of psychopathic disorder or mental illness showed that patients with a classification of psychopathic disorder were more likely to have been convicted of a standard list offence at any time after discharge ($\chi^2=4.5$, d.f.=1, $P=0.034$).

### Violent behaviour and fire-setting after discharge

For various reasons not every offence leads to a conviction – particularly when patients are detained in healthcare settings. Within the first 2 years after discharge at least 28% of patients had exhibited violent behaviour not resulting in conviction and within 5 years this had increased to at least 42%. Within the first 2 years after discharge at least 3% of patients had engaged in fire-setting not resulting in conviction; within 5 years this had increased to at least 6%.

### Readmission after discharge

The majority of patients in the sample were admitted to a psychiatric hospital at some point in the follow-up. Of the 151 patients who were discharged directly to the community, 40 (26%) were readmitted to a medium secure unit during the follow-up. A further 152 patients (27.6%) who were discharged to a hospital remained in hospital for the whole of the first year after discharge, including 40 patients transferred to high secure hospitals. Overall 20.5% were readmitted to Arnold Lodge, 7.8% to other medium secure units and 14.9% to a high security hospital, with 207 of the 550 discharged (37.6%) subsequently spending some time in medium or high security and with some patients spending time in both. Over the whole study period there were missing readmission data for 60 patients; excluding these, only 151 patients (30.8%) were never readmitted to a psychiatric hospital.

### Employment

On admission, 12.3% of patients had never been employed, and the majority (51.3%) had been employed in unskilled jobs. Stable employment after discharge was uncommon (14.5%) and was often provided by family members. Patients with a Mental Health Act classification of psychopathic disorder were more likely than patients with a classification of mental illness to have gained employment at some point during the follow-up period. This approached significance ($\chi^2=3.7$, d.f.=1, $P=0.054$).

### DISCUSSION

Medium secure units were introduced in the UK in the early 1980s to fill a perceived...
gap between local services and high secure care. Arnold Lodge was one of the first of these units to open, so an investigation of the course and outcome of its patients offers a unique opportunity to examine the long-term effectiveness of these expensive facilities. We found that after discharge the outcome for patients was poor, with a mortality rate that was six times that which one might expect; that almost half of those discharged had at least one reconviction; that almost two-thirds were readmitted within 5 years after discharge; and that their capacity to obtain and retain gainful employment was very limited. Subject to certain limitations that we shall now consider, these results ought to be a major cause for concern.

Limitations
The first major limitation was that this study examined the outcome from a single unit; with its unique policies and therapeutic ethos; hence its results may not be generalisable to all medium secure units. What is needed now is similar reports from other medium secure units with which these findings can be compared. The other major follow-up study (Maden et al, 1999) was based on the Dennis Hill Unit, which largely concentrated on rehabilitating patients from high security. Arnold Lodge itself was also unique among medium secure units in that it provided a dedicated treatment service for individuals with a diagnosis of personality disorder (McMurran et al, 1998). Although this is important in providing information on the course of such individuals after discharge, the impact of this group is unlikely to feature in other units. An additional limitation, as with any long-term follow-up, is that these outcome data reflect the therapeutic practice and policies of Arnold Lodge at the time; these have changed substantially over 20 years. Most importantly, the major factor influencing long-term outcome is likely to be the care patients received after discharge. This varied geographically across the region, from well-established community forensic services to poorly developed services relying on a series of locum consultants. There have also been major changes over time in treatment: for example the use of atypical antipsychotic medication and the availability of assertive outreach services, which might be expected to have a positive impact on an individual’s course, were only in evidence towards the end of the study.

However, these are criticisms that can be levelled at any long-term follow-up study (Stone, 1990).

While acknowledging these limitations, it is also important to point to a major advantage of the study. Section 60 approval meant that patient ascertainment was high, as patients’ consent to follow-up was not required. This is important in a group containing a large number of individuals with antisocial tendencies, where it is generally accepted that it is difficult to obtain such consent (Paris, 2003), and that if there is an ascertainment bias those who consent to be studied are likely to have the better outcome. The use of multiple data sources in our study also improved accuracy and reduced attrition (Friendship et al, 2001; Francis et al, 2002).

Mortality
One of the most striking findings of this study was that a six-fold increase in mortality compared with the general population compares unfavourably with rates reported for other psychiatric groups. For instance, Harris & Barracough (1998) in a major review reported SMRs of 156 and 141 respectively for men and women with schizophrenia and 184 for those with personality disorder. Similarly, a general psychiatry first admission sample (using a similar method of recruitment to our study) had an SMR of 136, which was not significantly raised compared with the general population, and no suicide, despite an almost complete follow-up over 16 years (Naik et al, 1997). A study of deaths in 1996–1997 among ex-prisoners, offenders on community sentences and prisoners reported SMRs of 276, 358 and 150 (Sattar, 2001). Thus, our sample had a significantly greater mortality than other psychiatric and criminal justice samples, and it may therefore be a genuinely ‘high-risk’ population in terms of suicide and unnatural death (for further discussion see Davies et al, 2001).

Death, particularly from suicide, represents the end-point of a number of complex and long-term processes. Following their first admission to a medium secure unit patients will have had a variety of experiences; many will have continued to receive psychiatric services; and some will have remained as in-patients over the entire period of follow-up. Some patients will have returned to the criminal justice system with no further psychiatric contact; others will have been discharged or lost to follow-up in the community. The vast majority will continue to experience mental disorder with its long-term risk of suicide and increased mortality. In addition these people will carry the stigma of previous offending, and many will be convicted of further offences; to this will be added further risk factors such as difficulties in obtaining employment, finding accommodation and maintaining social networks, resulting in poverty and social exclusion. Factors detrimental to physical health such as obesity, lack of exercise, smoking and the side-effects of antipsychotic medication (such as diabetes and cardiac arrhythmias) are also common in psychiatric populations.

The message for general psychiatric services that will in the main be responsible for such patients is that this population’s risks of mortality are high, probably related to psychiatric illness, treatment and lifestyle, and that all of these problems need to be addressed, as well as risks to others.

Reconviction
A methodological strength of our study was the use of multiple sources to minimise attrition and corroborate conviction data. Although the rates of conviction could be considered to be high, they are less than those found in other criminological samples. For instance, about a quarter (26%) of this sample were convicted of a standard list offence within 2 years of discharge (or 30% of those discharged directly to the community), compared with the 58% of prisoners released in 2001 who were reconvicted of a standard list offence within 2 years (Home Office, 2002). Reducing reoffending is difficult as the criminal justice system has discovered, with several initiatives and legislation only succeeding in reducing reoffending in England and Wales by 1.3% between 1997 and 2001 (Home Office, 2006). The lowest rates of reconviction are for those with the longest periods of detention and closest supervision, namely life-sentenced prisoners and restricted patients. The reconviction rates for standard list and grave offences at 5 years (for those with previous convictions) are 17% and 3% for restricted patients and 10% and 1% for life licences respectively (Kershaw et al, 1997). However, many discharged patients were also involved in violent incidents or acts of arson (42% and 6% respectively at 5 years) for which they were not charged or convicted.
Implications of the study

Overall the long-term outcome for former patients from the medium secure unit in this study was poor, with excess mortality, high rates of reconviction and readmission, and few gaining employment. Advances in mental health provision will, we hope, have a positive impact on an individual’s course in future years. For community services, particularly assertive outreach, community forensic and community mental health teams, the message must be that risk in terms of mortality, morbidity and harm to others remains high in this population over long periods. Follow-up care needs to be consistent and long-term, and information on risk should not be lost or overlooked – an increasingly difficult task with the multiplicity of teams and continual reorganisation of psychiatric services.

ACKNOWLEDGEMENTS

The initial description of the cohort (Ricketts et al, 2001) and follow-up using Office for National Statistics data were funded by Nottinghamshire Healthcare NHS Trust; further follow-up was funded by the NHS National Forensic Mental Health Research and Development Programme. The study took place at the Arnold Lodge medium secure unit; the continued support of this unit is gratefully acknowledged. We thank Dr Martin Tobin, Senior Lecturer in Genetic Epidemiology, University of Leicester, for advice on standardised mortality ratios, and Richard Twiner and Emma Booker for assisting in data collection.

REFERENCES


Impact of sexual violence on disclosure during Home Office interviews

DIANA BÖGNER, JANE HERLIHY and CHRIS R. BREWIN

Background Late disclosure or non-disclosure during Home Office interviews is commonly cited as a reason to doubt an asylum seeker’s credibility, but disclosure may be affected by other factors.

Aims To determine whether and how sexual violence affects asylum seekers’ disclosure of personal information during Home Office interviews.

Method Twenty-seven refugees and asylum seekers were interviewed using semi-structured interviews and self-report measures.

Results The majority of participants reported difficulties in disclosing. Those with a history of sexual violence reported more difficulties in disclosing personal information during Home Office interviews, were more likely to dissociate during these interviews and scored significantly higher on measures of post-traumatic stress symptoms and shame than those with a history of non-sexual violence.

Conclusions The results indicate the importance of shame, dissociation and psychopathology in disclosure and support the need for immigration procedures sensitive to these issues. Judgments that late disclosure is indicative of a fabricated asylum claim must take into account the possibility of factors related to sexual violence and the circumstances of the interview process itself.

Declaration of interest None. Funding detailed in Acknowledgements.

To be granted asylum under the 1951 United Nations Convention Relating to the Status of Refugees, the asylum applicant has to show a ‘well-founded fear of being persecuted in his or her country of origin for reasons of race, religion, nationality, membership of a particular social group, or political opinion’ (United Nations High Commissioner for Refugees, 1992). Since there is often little documentary evidence about the asylum seeker, credibility of the individual is key. Late disclosure, or description of incidents in later interviews of which no mention was made in the first, is commonly cited as a reason to doubt an asylum seeker’s credibility (see Asylum Aid, 1999). It is understandable that the addition of new evidence could be seen as evidence against the claimant’s honesty. However, this assumption may fail to take into account other reasons for not disclosing at the outset. To date, there has been no empirical study on what affects asylum seekers’ disclosure during legal interviews.

Many refugees who come to the UK have experienced or witnessed torture and organised violence (Burnett & Peel, 2001). Disclosure is specifically an issue with torture survivors owing to their difficulties of trust in other people (particularly those in authority) and their avoidance of painful memories (Medical Foundation for the Care of Victims of Torture, 2002). A meta-analysis revealed increased prevalence rates of post-traumatic stress disorder (PTSD) in refugees resettled in Western countries (Fazel et al., 2005). Symptoms of PTSD may be activated during the Home Office interview as a result of being reminded of the traumatic event, which in turn might reduce a person’s ability to give a coherent account and might lead to non-disclosure.

There is also evidence that different trauma types are associated with different PTSD patterns. Two studies found a significant relationship between sexual torture and the avoidance criteria of PTSD (Ramsey et al., 1993; Van Velsen et al., 1996). Van Velsen et al. (1996) speculated that the intimate nature of the sexual attack and associated negative emotions, such as feelings of humiliation and shame, are likely to be critical elements leading to subsequent avoidance behaviour. However, this has not been specifically tested.

Refugees and asylum seekers often come from cultures with different attitudes towards sexuality. Sexual violence and rape are often taboo subjects and can bring about feelings of shame. Women who have been subjected to sexual assault may be shunned by their community and family if they admit to this and therefore may not disclose it in their asylum interview (United Nations, 1997; Burnett, 1999). Men also tend to underreport experiences of sexual violence (Peel et al., 2000). Feelings of shame have been mentioned in the literature as a factor affecting disclosure (Hill et al., 1993) and there have been several empirical studies demonstrating the relationship between shame and disclosure (Swan & Andrews, 2003; Hook & Andrews, 2003). There is also increasing evidence that shame may be linked to the course or onset of PTSD (Andrews et al., 2000; Leskela et al., 2002).

The study of different trauma types by Van Velsen et al. (1996) suggested including a measure of dissociative phenomena in future research, as dissociation might be closely related to PTSD avoidance symptoms. Indeed, dissociative experiences are commonly reported by individuals with a diagnosis of PTSD (Ozer et al., 2003). Carlson & Rosser-Hogan (1991) found high levels of association between traumatic experiences and the severity of both traumatic stress and dissociative reactions in a group of 50 Cambodian refugees. However, dissociative responses not only occur as an aftermath of a traumatic event, but can also be experienced at the time of the trauma (peritraumatically; Weiss et al., 1995). Dissociative reactions might be activated during an anxiety-provoking event, such as the Home Office interview, which might affect disclosure.

The first aim of our study was to investigate the impact of sexual violence on refugees’ and asylum seekers’ reported post-traumatic stress symptoms, shame reactions, dissociative experiences and difficulties in disclosure during Home Office interviews. The second aim of the study was to explore more systematically the factors involved in refugees’ and asylum
seekers’ disclosure during Home Office interviews by means of a qualitative semi-structured interview.

**METHOD**

**Sample and procedure**

Refugees and asylum seekers with a history of pre-migration trauma were included in the study. Twenty-seven participants in total were recruited from a central London traumatic stress clinic (n=17) and two London-based community services (n=10). They were invited to take part in a research study about refugees’ and asylum seekers’ experiences of legal interviews; demographic data are reported in Table 1. The participants, who had arrived in the UK between 1995 and 2003, originated from 14 countries in Europe, Africa, the Middle East and Latin America. Written informed consent was obtained.

At the time of testing, 15 of the 27 study participants were receiving psychological input at a specialist tier 3 London traumatic stress clinic. Nine of them were receiving long-term weekly individual psychological treatment, and 6 had just completed a 3-month weekly psychoeducation group. This group was run for people who were on the waiting list for individual psychology. The remainder (n=12) had not received psychological input since coming to the UK.

Participants were divided into two groups. The first group consisted of participants with a history of sexual violence. Following the study by Van Velsen et al. (1996), sexual violence was defined as rape (of men or women) or other tortures directed to the genital area. The second group consisted of participants with a history of non-sexual violence. This was broadly defined as having experienced or witnessed some form of psychological and/or physical maltreatment including torture. Overall, 15 participants experienced some form of sexual violence, including rape (n=12) and sexual torture (n=3). Twelve participants experienced or witnessed some other form of violence, including torture (n=6), being shot (n=2), beatings (n=2) and witnessing killing of family members (n=2). This information was obtained, with consent, from the person’s clinician or caseworker, or from medical notes. All participants had had a screening interview shortly after their arrival in the UK, followed by one or more main Home Office interviews; 24 participants had had one main Home Office interview and 3 participants had had two main Home Office interviews.

Research interviews took place over a 6-month period from November 2004 to May 2005. Participants were interviewed on one occasion about their main Home Office interview. People who had attended two main Home Office interviews were questioned about their first one. Interpreters were used when requested by participants. Seven participants were interviewed with the assistance of an interpreter who was officially accredited. To avoid translation issues, all measures were presented orally during the interview.

**Measures**

**PTSD Symptom Scale—Interview**

The PTSD Symptom Scale—Interview (PSS-I; Foa et al., 1993) was used to assess current PTSD symptoms according to DSM–IV criteria (American Psychiatric Association, 1994). This is a semi-structured interview consisting of 17 items; answers are rated from 0 (not at all) to 3 (five or more times per week/very much). Total severity scores are based on the sums of the raw items.

**Hopkins Symptom Checklist**

The Hopkins Symptom Checklist–25 (HSCL–25; Derogatis et al., 1974) was included since depression has been found to be highly comorbid with PTSD (Blanchard et al., 1998). The HSCL–25 was chosen for its cross-cultural robustness (Kinzie & Manson, 1987). Participants completed part 2 of the scale, which has 15 depression items rated on a four-point scale, ranging from 1 (not at all) to 4 (extremely). The mean of the 15 depression items has been shown to correlate with major depression as defined by the DSM–IV (American Psychiatric Association, 1994).

**Experience of Shame Scale**

The Experience of Shame Scale (ESS; Andrews et al., 2002) is a 25-item scale assessing three different domains of shame: characterological, behavioural and bodily shame. Within each of these domains there are items reflecting the experiential (feeling shame), cognitive (concern over others’ opinions) and behavioural (concealment or avoidance) components of shame. Participants rate each item according to how they have felt in the past year, on a four-point scale ranging from 1 (not at all) to 4 (very much).

**Peritraumatic Dissociative Experiences Questionnaire**

The Peritraumatic Dissociative Experiences Questionnaire—Self-Report Version (PDEQ–SRV; Marmar et al., 1997) consists of ten items measuring retrospectively acute dissociative reactions during a specific event. Items are rated on a five-point scale, ranging from 1 (not at all true) to 5 (extremely true). Participants were instructed to complete the items based on their experiences and reactions during the Home Office interview and immediately afterwards.

**Difficulty in disclosure**

Participants were asked to rate on a four-point scale, ranging from 1 (not at all) to 4 (extremely), how difficult they found it to disclose personal information during the Home Office interview.

**Semi-structured interview**

A semi-structured interview was used to collect qualitative data regarding people’s disclosure during Home Office interviews. Because of ethical constraints we did not set out to investigate whether sexual victimisation was disclosed or not during the Home Office interviews. Interviews were taped and transcribed. Four participants did not want their interview recorded and in these cases process notes were taken instead. Participants were asked a number of general questions relating to the disclosure of their index trauma.

(a) When was the first time you talked about what happened to you in (your home country)? After the event? After your arrival in the UK?

(b) Who did you talk to?

(c) Was there anything you initially did not tell this person?

Other questions specifically concerned disclosure behaviour during the Home Office interview.

(d) To what extent did you feel you could open up and talk openly about what happened?

(e) Are there any things you have not yet told the Home Office about? If yes, could you tell me what some of the reasons might be that you have found it difficult to do that?
Finally, a question was included to assess whether participants could identify any aspects relating to their cultural background that had affected disclosure during their Home Office interview, because research has shown that issues such as sexual violence are not readily disclosed to others owing to feelings of shame, social stigma and the risk of being shunned by family members and the community.

(f) Are there things you have not talked about because in your culture it is considered wrong?

Other questions assessed participants’ experiences of the Home Office interview, particularly addressing interpersonal and situation- and context-specific factors, as well as other issues and recommendations. These data will be reported separately.

Demographic and clinical factors

Demographic data were collected for all participants, including age, gender, nationality, current asylum status, dates of arrival in the UK, number and dates of Home Office interviews, decision on asylum claim following Home Office interview, time elapsed between Home Office interview and research interview (in months), and receipt of psychological treatments since arrival in the UK.

Statistical analysis

Several variables had skewed distributions and required transformation. Following Tabachnick & Fidell (2001), analyses using untransformed data are reported, as transformation did not affect the results. Differences between sexual and non-sexual violence groups on demographic factors and measures of PTSD, depression, shame, dissociation, and difficulty in disclosure were investigated using independent t-tests. Analysis of covariance was used to control separately for the effects of relevant variables on group differences in difficulty of disclosure. Correlations between age, time lag between Home Office and research interviews, PTSD, depression, shame, dissociation and difficulty in disclosure were examined using Spearman’s rho. Partial correlations were used to determine whether the associations between dissociation and shame and dissociation and disclosure were still significant after total PTSD symptoms were controlled for. Independent t-test was used to measure the relationship between difficulty in disclosure and decision on asylum claim, as well as receipt of psychological treatments. Statistical analyses used the Statistical Package for the Social Sciences version 11.5 for Windows. A two-tailed α level of \( P=0.05 \) was used to determine statistical significance.

The qualitative data were analysed using a thematic analysis approach, which focuses on identifiable themes and patterns of personal experiences (Aronson, 1994). Following recommendations by Elliot et al (1999), credibility checks were provided in several ways. To provide checks on reliability, a second marker audited the data from each question, looking at the themes created. Any differences in opinion were discussed and rectified. Furthermore, the findings were triangulated by comparing the outcome of the qualitative data with the results of the quantitative data and drawing parallels between the two (see Discussion). The validity of the conclusions drawn from the interview data is enhanced in several ways: first, we present direct quotes from the interviews to demonstrate to the reader the relationship between themes and the source data; second, to indicate how representative the themes were of the sample as a whole, the proportion of participants for each theme is outlined; and third, the analysis includes a negative case analysis, which means reporting on minority as well as majority responses.

RESULTS

Quantitative findings

No significant group difference existed for age, time lag in months between participants’ main Home Office interview and research interview, PTSD re-experiencing symptoms, PTSD arousal symptoms or depression (Table 1). Those with a history

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Comparison of groups by measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sexual violence</td>
</tr>
<tr>
<td>Gender, n</td>
<td>(n=15)</td>
</tr>
<tr>
<td>Male</td>
<td>4</td>
</tr>
<tr>
<td>Female</td>
<td>11</td>
</tr>
<tr>
<td>Asylum status, n</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>6</td>
</tr>
<tr>
<td>ELR</td>
<td>2</td>
</tr>
<tr>
<td>Under appeal</td>
<td>7</td>
</tr>
<tr>
<td>Asylum decision following Home Office interview, n</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>6</td>
</tr>
<tr>
<td>No</td>
<td>9</td>
</tr>
<tr>
<td>Psychological treatment, n</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>Age, years: mean (s.d.)</td>
<td>37 (12.0)</td>
</tr>
<tr>
<td>Time lag in months: mean (s.d.)</td>
<td>40.5 (25.3)</td>
</tr>
<tr>
<td>Scores: mean (s.d.)</td>
<td></td>
</tr>
<tr>
<td>PSS-I</td>
<td></td>
</tr>
<tr>
<td>Overall severity</td>
<td>37.7 (10.7)</td>
</tr>
<tr>
<td>Re-experiencing</td>
<td>8.9 (3.0)</td>
</tr>
<tr>
<td>Avoidance</td>
<td>16.0 (4.4)</td>
</tr>
<tr>
<td>Hyperarousal</td>
<td>12.7 (4.5)</td>
</tr>
<tr>
<td>HSCL depression</td>
<td>43.5 (11.4)</td>
</tr>
<tr>
<td>ESS</td>
<td>65.6 (19.6)</td>
</tr>
<tr>
<td>PDEQ–SRV</td>
<td>31.9 (10.1)</td>
</tr>
<tr>
<td>Difficulty in disclosure</td>
<td>3.5 (0.9)</td>
</tr>
</tbody>
</table>

ELR, exceptional leave to remain; ESS, Experience of Shame Scale; HSCL, Hopkins Symptom Checklist; ILR, indefinite leave to remain; PDEQ–SRV, Peritraumatic Dissociative Experiences Questionnaire – Self-Report Version; PSS-I, PTSD Symptom Scale – Interview.
of sexual violence reported greater overall PTSD severity and avoidance symptoms, as well as greater feelings of shame (Table 1). This group also described more dissociation symptoms and greater difficulty in disclosure of personal information during their Home Office interview.

There was no association between self-disclosure and decision on asylum claim (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure ((and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclosure (and disclose:)

Discllosure behaviour during Home Office interview

Three different themes emerged from the participants’ answers: no reported problem in opening up; finding it too difficult to disclose; and wanting to disclose, but not being given the chance to do so. Seven people reported no difficulties with opening up and disclosing personal details in their Home Office interview. Twelve people reported difficulties in disclosing personal details during the Home Office interview; 10 of them had a history of sexual violence. Reasons cited were feeling too traumatised, afraid and ashamed to talk about the past (n=10), which resulted in them not being able to tell the Home Office interviewer what had happened to them or to answer questions.

It was the first time in my life that I had to talk about what happened to me. I only told the interviewer about 10%, I could not talk, it was too difficult. I felt so traumatised and ashamed. (P2)

Further reasons cited were intrusive experiences, such as intrusive memories and flashbacks, which affected their ability to focus on the interview and give a coherent account (n=2):

When I talked about the past, what happened to me, the memories came, flashbacks. And then I found it difficult to remember anything that happened in my country. I was crying, I was shocked. It was hard to explain what happened to me. (P9)

Others reported dissociative experiences that made it difficult for them to focus on the interview, and affected their ability to disclose:

I tried to talk, but my mind kept wandering off and I kept thinking about the trauma and my family that I lost. Everything seemed unreal to me, I felt like I was dreaming. I found it hard to focus on the interview and answer questions. (P6)

Ten people reported that they wanted to tell the Home Office what had happened to them, but that they were not given the...
opportunity to do so; the interviewer was apparently more interested in factual details about their home country and how they got to the UK than what had happened to them or their families:

'I wanted to explain properly, but they just stopped me. They ask you to make it short and give yes or no answers. You don’t get a chance to say much or explain to them. Therefore I did not go into much detail. But that affected me later [at the court] when I was asked why I did not tell them in the [Home Office] interview.’ (P15)

Five of the people who wanted to disclose also reported that they were asked similar questions repeatedly, which increased their stress levels and affected on their ability to disclose:

‘When he asked me questions and I answered them he started cross-examining me. The more I said the more questions he asked me. It felt like he was trying to trick me. I felt nervous and stressed, which made it harder to talk for me.’

(P16)

Fifteen people reported that there are still things they have not told the Home Office about; 10 were men and women with a history of sexual torture, and most of them reported feelings of shame as a reason for non-disclosure (n=7):

‘I wanted to keep things from my past private. I was scared that they would look at me badly and make me feel ashamed. I could not tell everything at the interview, but later on I was able to tell the court. They were nice at the court and made me feel more relaxed.’

(P21)

Other reasons included forgetting some details, which they were not able to mention subsequently in later interviews for fear it would affect their credibility (n=2); being unsure whether they could disclose details they were not directly asked about (n=3); and not being given the opportunity and the time to talk openly about their past traumatic experiences (n=2).

Cultural factors affecting disclosure

Eight participants reported that there were things they have not talked about because in their culture it is considered wrong; all of them were men and women with a history of sexual violence. Most of them stated that in their culture sexual issues are not talked about, especially rape:

‘At home you are not allowed to talk to other men you are not related to, you are not allowed to look any men in the eyes. So how could I have looked him [male Home Office official] in the eyes and told him what happened to me – it’s a different culture.’

(P11)

Two individuals specifically mentioned feelings of shame associated with rape, and that shame had prevented them from talking about the rape in the interview:

‘There is a lot of shame associated with what I experienced. Shame in my culture prevents me from talking about this.’

(P17)

Direct disclosure of sexual victimisation

Although data on disclosure of sexual victimisation were not specifically collected, further analysis of the transcripts revealed that of the 15 people with a history of sexual violence, 5 reported that they had disclosed sexual victimisation, including rape, during their Home Office interview, and 6 did not disclose it. It is unclear whether the remaining 4 specifically disclosed sexual victimisation. Interestingly, everybody who disclosed a history of sexual violence reported being prevented from talking about it further in the interview by the Home Office official.

DISCUSSION

This study refined and extended previous findings by Van Velsen et al (1996) by demonstrating that there is a significant association between shame and PTSD avoidance symptoms, which suggests that shame might act as a mediator between a history of sexual violence and PTSD avoidance symptoms. Shame was also significantly associated with overall PTSD severity, which provides further evidence that shame might be linked to the course and onset of PTSD (Andrews et al, 2000; Leskela et al, 2002). The significant relationship between dissociation and PTSD avoidance symptoms confirms speculations by Van Velsen et al (1996). The results are also in line with research showing that dissociative experiences are commonly reported by individuals with a diagnosis of PTSD (Ozer et al, 2003). Furthermore, our analysis revealed that those who experienced higher levels of dissociative experiences during the Home Office interviews were those who had higher levels of shame.

Data from the qualitative interviews provide further evidence for the above findings. Perhaps one of the most striking findings was that 20 participants talked for the first time about their pre-migration trauma only after entering the UK, and of those, 13 talked to Home Office officials. These findings underscore the degree of avoidance associated with the experience of trauma and are likely to be very relevant to the large numbers of refugees coming to the UK who have experienced or witnessed torture and organised violence (Burnett & Peel, 2001).

Many participants reported difficulties with disclosing personal details in their Home Office interview, and reasons frequently cited for this were negative emotions such as feeling too traumatised by past experiences or feelings of shame. Shame was especially salient for people with a history of sexual violence. Many of those reported that in their culture sexual issues are not discussed with others, and that this prevented them from disclosing sexual issues during their Home Office interview. This supports previous findings that shame is associated with difficulty in disclosure (Swan & Andrews, 2003; Hook & Andrews, 2003) and is consistent with Hill et al (1993) who found that sexual issues often remain too shameful to discuss, even in therapy.

Participants also reported experiencing psychological symptoms during Home Office interviews, such as dissociative experiences, flashbacks and avoidance behaviours (e.g. avoiding thoughts or feelings associated with the trauma and not being able to remember details), which had an impact on their ability to disclose. This suggests that people’s psychological states should be routinely evaluated when assessing their ability to give a coherent personal history in an interview with officials.

Finally, it should be noted that although the difficulties with disclosure seemed to be persistent, many participants did express a willingness to talk to officials about their experiences. However, some described not being given the opportunity to do so or being prevented by the interviewer from discussing their experiences. One explanation could be vicarious traumatisation of the interviewers, which is a common phenomenon in people working with trauma survivors (Figley, 1995). Indeed, a multidisciplinary analysis of the decision-making process of the Canadian Immigration and Refugee Board showed that coping with vicarious traumatisation and uncontrolled emotional reactions was one of the factors having a negative impact on the board members’ ability to evaluate credibility and on the overall conduct of hearings (Rousseau et al, 2002). This needs to be clarified by further research.

In summary, our results indicate that late disclosure or non-disclosure during Home Office interviews does not necessarily imply a lack of honesty on the asylum
seeker’s part, and highlight that disclosure is complex and influenced by a variety of factors that need to be taken into account when judging asylum seekers’ credibility based on the information they disclose. A Home Office interview can be a stressful and anxiety-provoking event, which may provoke reactions that interfere with disclosure.

**Limitations of the study**
Several methodological aspects of this study warrant consideration. The sample size was small, which ruled out the use of multivariate analyses. Language and cultural barriers presented an obstacle, as they made the collection of accurate data more difficult and might have increased measurement error. There is also the potential for a sampling bias, especially finding a sample that is representative of the general refugee and asylum-seeker population. However, this is an applied study of a real life situation, representing the diverse population of refugees going through asylum interviews in the UK. Van Velsen et al. (1996) suggested that sampling biases generally pose a problem in research studies on refugees and asylum seekers, as this population is already exposed to numerous selection biases. Similarly, the lack of a control group restricts our findings. The comparisons are limited because the base rates of PTSD, shame, depression, dissociation and difficulty in disclosure are unknown in this group. It would, of course, be desirable to find a comparison group of refugees and asylum seekers who had not experienced any kind of violence. Whether there are refugees and asylum seekers who fit these criteria depends largely on the definition of violence and the definition of ‘refugee’ itself. None the less, the above issues restrict the generalisability of the findings and the tentative conclusions outlined in this paper should be considered with this in mind.

Another limitation concerns people’s accuracy in reporting emotional experiences that occurred several months or even years ago. However, since there is no significant difference between groups in the length of time between Home Office interviews and research interviews, this is unlikely to affect the interpretation of the data significantly. On a similar note, dissociation may in some cases have been experienced after the interview. Finally, the cumulative effect of multiple traumas needs to be considered; the greater difficulty in disclosure in the sexual violence group may be related to the fact that some people with a history of sexual violence also experienced physical trauma.

**Implications of our findings**
The above findings have implications for the process of granting asylum in the UK. Asylum seekers often come from countries where they experienced or witnessed torture and organised violence, which means that they are in a vulnerable position when entering the UK. Most asylum seekers in our study experienced the immigration process – including the Home Office interviews – as stressful and anxiety-provoking, because many feared deportation. Disclosure is a difficult issue in this group; many need time to process past traumatic events and to establish a sufficient level of trust and confidence to reveal the potentially painful and shaming details of their experiences. This needs to be taken into account by an immigration system that requires asylum seekers to make a claim shortly after arrival. It is therefore of paramount importance that sensitivity is used when processing refugee claims and that immigration officials are aware of the needs of asylum seekers in order to avoid inducing further distress in this already highly traumatised group.

The findings also have implications for current immigration policy. The need for policies that identify asylum seekers who fabricate their stories and that deter immigrants who have left their country for economic reasons seems understandable. However, this study suggests that legitimate asylum seekers may be punished and retraumatised by the enforcement of some of these policies. Furthermore, the immigration system needs to take into account the special needs of victims of sexual violence, particularly since there is a high incidence of shame in this group. Given the significant associations between shame, PTSD avoidance symptoms and difficulty in disclosure, one might speculate that being forced to talk about a traumatic event could potentially activate shame reactions, and that people experiencing more shame are engaging in strategies to avoid this feeling, such as non-disclosure of sensitive personal information. This also highlights the importance of recognising and dealing with asylum seekers’ shame in an empathic way. It seems that immigration officials could benefit from supervision and training in how to recognise stress reactions in interviewees, such as PTSD symptoms, shame and dissociative experiences, as well as an awareness of the impact of these on people’s psychological health, affective states and ability to disclose.

**ACKNOWLEDGEMENTS**
The University College London Central Research Fund is gratefuly acknowledged for funding this study. We also thank Dr Stuart Turner for his very helpful comments on this paper. This work was undertaken with the support of Camden and Islington Mental Health and Social Care Trust, who received funding from the NHS Executive.

**REFERENCES**


Influence of drug company authorship and sponsorship on drug trial outcomes

TONGEJI TUNGARAZA and ROB POOLE

Summary
Studies of drug treatments are more likely to report favourable outcomes when they are funded by the pharmaceutical industry. We compared drug trials reported in three major psychiatric journals to investigate these influences. Independent studies were more likely to report negative findings than industry-funded studies. However, the involvement of a drug company employee had a much greater effect on study outcome than financial sponsorship alone.

Declaration of interest
T. T. has accepted sponsorship to attend conferences from Janssen-Cilag and Eli Lilly. R. P. has accepted speakers’ fees from Lundbeck, Eli Lilly and Pfizer; he has also accepted sponsorship to attend conferences from Wyeth, AstraZeneca and Eli Lilly. No funding was obtained for the present study.

It is known that studies of drug treatments are more likely to report favourable outcomes when they are funded by drug companies (Bekelman et al., 2003; Lexchin et al., 2003). There is also concern over the conflicts of interest created by authors’ personal financial links to companies (Bodenheimer, 2000; Komsaroff & Kerridge, 2002). Most studies of these influences are based upon randomised controlled trials in internal medicine. The study reported here concerns a broad range of drug trials in psychiatry. We explore the difference between having an author who is an ‘employee’ of a drug company (defined here as holding a consultancy, being an employee or being a shareholder) and receiving financial support from a drug company, and how these influence study outcome in comparison with independent studies.

METHOD
The British Journal of Psychiatry, American Journal of Psychiatry and Archives of General Psychiatry were selected as being widely read journals. They were surveyed for original data-based papers concerning psychiatric drug treatment, published between January 2000 and December 2004 inclusive. All methodologies were included (e.g. randomised controlled trials (RCTs), studies of drug levels in breast milk). Journals were searched manually and information was collected from full-text versions.

Outcomes were rated by T. T. He was aware of funding, as this was apparent in the papers. Studies were classified as reporting positive findings if they clearly stated that use of the index drug led to a better clinical outcome or was better tolerated than another treatment. Studies were classified as reporting negative findings if they clearly stated that use of a comparison treatment led to a better outcome or was better tolerated than the index drug or that there was no difference in clinical outcome or tolerability. Where the conclusions in the full text and abstract were equivocal, T. T. made a judgement as to whether the balance of findings was positive or negative.

Papers were included from all psychiatric sub-specialities. Outcome studies were included that compared an index drug with placebo, another drug or a psychological therapy. Studies were excluded if they concerned an index drug that was long established (e.g. tricyclic antidepressants, lithium, older antipsychotics) unless it was being used for a novel indication (e.g. testosterone for resistant depression). Short reports, letters to the editor, editorials, review articles and meta-analyses were excluded.

The authors’ relationship with the drug company was determined from declared affiliations and conflicts of interest, or from acknowledgements. Studies were classified as industry-funded if the study was wholly or partly funded by a drug company, including funding in kind (provision of drugs and placebos, or an author who was an employee). Authors were regarded as employees if they worked full time for the company, or declared consultancy positions or shareholdings. Studies were regarded as independently funded if sufficient information was provided to exclude any of these relationships.

RESULTS
Of the 198 studies that met the inclusion criteria, 8 (4%) lacked sufficient information on funding and were excluded. The remaining studies fell into three groups:

(a) studies funded independently of the drug industry (‘independent’);

(b) studies with one or more authors employed by a drug company (‘industry-authored’);

(c) studies funded by industry but without an employee author (‘industry-sponsored’).

Of these 190 studies, 33 (17%) were published in the British Journal of Psychiatry, 98 (52%) in the American Journal of Psychiatry and 59 (31%) in the Archives of General Psychiatry. Most studies (157) concerned adults; the remainder concerned elderly people, children, or mothers and babies. Of the 132 studies that were randomised controlled trials, 112 (85%) were industry-funded. In 75% of studies the index drug was an antipsychotic or an antidepressant (Table 1).

There was a significant difference between journals in reporting of negative results, the British Journal of Psychiatry being more likely to report negative findings than the other two ($\chi^2=7.99$, d.f.=2, $P=0.0184$).

Financial relationship with the drug industry
Forty-four studies (23%) were independent. Of the 146 that were industry-funded, 58 (40%) also received funding from a non-industry source. Six pharmaceutical companies funded nearly half of all the studies surveyed. There were 76 industry-authored studies (40%); of these, 64 (84%) had authors who were employees or shareholders. Seventy studies (37%) were industry-sponsored.
Outcomes

Positive findings were reported in 152 (80%) studies, whereas 38 (20%) reported negative findings. Independent studies were more likely to report negative findings than industry-funded studies. Sixteen (36%) of the 44 independent studies reported negative findings compared with 22 (15%) of the industry-funded studies. The difference was statistically significant (Yates' corrected $\chi^2=8.3$, d.f.=1, $P=0.004$). Only two (3%) of the 76 industry-authored studies reported negative findings. The difference between this group and the independent studies was highly statistically significant (Yates’ corrected $\chi^2=22.29$, d.f.=1, $P<0.0001$). A similar statistically significant difference was observed in the reporting of negative findings between industry-authored and industry-sponsored studies (Yates’ corrected $\chi^2=17.18$, d.f.=1, $P<0.0001$). There was no significant difference between independent and industry-sponsored studies in reporting of positive or negative findings ($\chi^2=0.44$, d.f.=1, $P=0.51$).

DISCUSSION

The involvement of a drug company employee seems to exert a powerful effect on study outcome, whereas merely accepting industry sponsorship appears to have little or no effect. This finding is both novel and counter-intuitive. One might expect that the difference between the two forms of industry funding would be subtle. In fact, the difference is highly statistically significant, in contrast to the lack of difference between studies with financial sponsorship only and fully independent studies.

There are some factors that might have confounded our findings. There were more RCTs among the industry-funded studies. Unlike other investigators, we included all methodologies because the number of independent RCTs in psychiatry is small. It might be that RCTs are intrinsically more likely to produce positive findings. Equally, they might be particularly vulnerable to being abandoned when preliminary findings are not promising (Henry et al, 2005). We did not assess the scientific quality of different studies. It is possible that independent studies tend to be statistically underpowered and that this leads to overreporting of negative findings (Djulbegovic et al, 2000; Procyshyn et al, 2004).

Our findings are unlikely to be solely due to these factors. All previous studies comparing industry-funded RCTs with independent ones have shown that the former are more likely to report positive findings. If industry-funded studies are less likely to be underpowered or methodologically flawed, then one would expect that the reporting of negative findings would be similar in the industry-authored and industry-sponsored groups, whereas actually the sponsored and independent studies were similar. We seem to have found an ‘all or nothing’ effect related to the involvement of a drug company employee.

In conclusion, we have confirmed previous findings that industry-funded studies are less likely to report negative findings. Our novel finding is that this effect appears to be largely or exclusively due to the presence of a company employee among the authorship. This finding requires replication with attention to differences in studies’ methodological rigour and statistical power, in order to exclude these as confounding variables.

REFERENCES


Depressive symptoms during pregnancy and low birth weight at term
Longitudinal study

JONATHAN EVANS, JON HERON, ROSHNI R. PATEL and NICOLA WILES

Summary
There is conflicting evidence regarding the effect of depression during pregnancy on birth weight. We used data from the Avon Longitudinal Study of Parents and Children to investigate whether depressive symptoms during pregnancy in 10,967 women led to low birth weight at term in their offspring. Those with a high depressive symptom score during pregnancy were more likely to have babies of low birth weight (95% CI 1.16–2.40, P < 0.01), but this attenuated after adjustment for confounders (OR = 1.29, 95% CI 0.87–1.91, P = 0.210). Hence there is little evidence of an independent association between depressive symptoms during pregnancy and birth weight.

Declaration of interest
None.

Low birth weight at term is a marker of intrauterine growth restriction and can result from prenatal exposure to high maternal levels of steroids (Gur et al., 2004). As maternal stress results in raised cortisol, psychological symptoms during pregnancy may cause intrauterine growth restriction. Depression and anxiety are common during pregnancy, but evidence for any effect on intrauterine growth is conflicting (e.g. Kelly et al., 2002; Andersson et al., 2004). In this large prospective study we investigated whether mothers with symptoms of depression and anxiety during the second trimester of pregnancy are at greater risk of having babies of low birth weight at term.

METHOD

We used data collected as part of the Avon Longitudinal Study of Parents and Children (ALSPAC; http://www.alspac.bris.ac.uk). In total, 14,541 women were enrolled and 13,194 (90.7%) completed a questionnaire at 18 weeks’ gestation.

Questionnaires included the Edinburgh Postnatal Depression Scale (EPDS; Cox et al., 1987) and sub-scales of the Crown–Crisp Experiential Index (CCEI), which measures ‘free-floating’ anxiety (Crisp et al., 1978) which were completed at 18 and 32 weeks of pregnancy. The EPDS has been validated for use during pregnancy and at other times (Murray & Cox, 1990; Shakespeare, 2001). The CCEI correlates well with other measures of anxiety in women during pregnancy and the internal consistency exceeds 0.8.

Birth weight was obtained directly by research staff. We included only those babies delivered between 37 and 43 weeks’ inclusive and we excluded multiple births as these babies are more likely to be of lower birth weight.

Data on confounding factors came from the self-report postal questionnaires and included gender of the baby, gestational age, smoking, maternal age, parity, educational level, alcohol and caffeine use during pregnancy, chronic disease, pre-pregnancy body mass index, ethnicity and a history of miscarriage, Caesarean section, prior low birth weight, or preterm births.

We used linear and logistic regression analyses with birth weight as a continuous outcome or dichotomised as low birth weight (<2500 g) at term. We used the multiple imputation by chained equation method to impute missing data for confounders (van Buuren et al., 1999).

RESULTS

Analyses were performed on 10,967 women, of whom 1519 (13.9%, 95% CI 13.2–14.5%) scored >12 on the EPDS at 18 weeks of pregnancy, indicating a likely diagnosis of depression. There were 8427 (76.8%) women with complete data on all confounding variables. We imputed missing data on confounders in all adjusted analyses.

Women scoring >12 on the EPDS at 18 weeks of pregnancy had babies that were 33.8 g lighter (95% CI 8.0–59.7, P = 0.01) compared with women with an EPDS score <13. Following adjustment for confounders, an EPDS score >12 was no longer significantly associated with birth weight: babies born to mothers with probable depression were 1.1 g lighter (95% CI −22.1 to 24.3, P = 0.4). The strongest confounder appeared to be smoking. The coefficient following adjustment for gender, maternal age and gestation was −25.9 (95% CI −49.9 to −2.02), this dropped to −1.0 (95% CI −24.9 to 22.9) following additional adjustment for smoking. Using anxiety as the exposure variable we found similar results (details available from the authors).

Following adjustment, there was no effect restricted to those with higher symptom scores: mothers scoring >16 on the EPDS at 18 weeks of pregnancy had babies that were 22.5 g lighter (95% CI −12.5 to 57.4, P = 0.207) than those scoring <13.

We repeated these analyses with the outcome categorised into low birth weight at term according to the traditional definition of <2500 g (Fig. 1). The drop in the estimated odds ratio was greatest following adjustment for smoking. The odds ratio following adjustment for gender, gestation and maternal age was 1.57 (95% CI 1.08–2.29), dropping to 1.38 (95% CI 0.94–2.01) after adjustment for smoking.

With anxiety as the exposure variable we found very similar results to those seen for depression (details available from the authors). We found a weak association with prolonged exposure to depressive symptoms. Those scoring >12 on the EPDS at both 18 and 32 weeks of pregnancy had babies that were 40.2 g lighter (95% CI 3.8–76.6, P < 0.05) than those with an EPDS score <13 at both times. This association diminished and was no longer statistically significant following adjustment.

DISCUSSION

This is the largest longitudinal study investigating the association between maternal mood during pregnancy and low birth weight. We found no evidence to support an independent association between depressive or anxiety symptoms during pregnancy and low birth weight at term. There is a widespread assumption that depression and anxiety during pregnancy can impair...
We addressed this problem by restricting our analyses to babies born at term.

In this study we adjusted for a range of confounding variables including smoking. It remains possible that smoking lies on the causal pathway between depression and intrauterine growth restriction. Adjustment for smoking removed any evidence of a direct association between mean birth weight at term and mood during pregnancy.

Although a woman with a high depressive symptom score in the mid trimester of pregnancy has a slightly increased risk of having a low birth weight infant, this is most likely to be owing to health behaviours associated with depressive symptoms.

Low birth weight at term is only one marker of foetal development. Studies in animals suggest there are long-term effects of exposure to prenatal stress (Barbazanges et al., 1996). It is possible that psychological symptoms during pregnancy have other effects on development of offspring that only become apparent after birth. This remains an important area for investigation (O’Connor et al., 2002).

REFERENCES


Aripiprazole – data on efficacy and associated mortality

El-Sayeh et al (2006) raise some important issues regarding the design and reporting of clinical trials. However, we feel that the conclusion that ‘ariipiprazole has been licensed despite the fact that few reliable data on this drug are publicly available’ merits further clarification. Aripiprazole was first approved in November 2002 in the USA, and in 2004 in Europe, based on the submission of a substantial body of evidence to the regulatory authorities on more than 4000 patients. However, Bristol-Myers Squibb and Otsuka Pharmaceuticals are committed to reporting trial results as completely as possible, and publication of pivotal studies has taken place subsequent to approval.

All the aripiprazole clinical studies were conducted in accordance with regulatory requirements and using accepted standards (Marder et al, 2003; Naber & Lambert, 2004). Such studies have inherent restrictions, and we recognise that patients enrolled may not always reflect those seen in everyday care. We understand the value of all study types – randomised controlled trials, naturalistic, retrospective, observational – in helping to determine the benefit-risk profile, and have recently completed a series of studies with more naturalistic designs and with large sample sizes, to explore the benefits in a wide range of patients (Tandon et al, 2006; Kerwin et al, 2007, details of the other study can be obtained from http://www.clinicaltrials.gov, trial number NCT00237939). These complete studies support the profile of aripiprazole established in the clinical studies reviewed by El-Sayeh et al in their systematic analysis.

With respect to the suggestion that deaths occurring during the aripiprazole studies have not been widely reported, it is our practice to report any deaths or adverse events applicable to a study and we have done so consistently in our publications. Deaths unfortunately do occur during studies, just as they do in real-world situations.

We are committed to continued openness and disclosure of clinical study results, and as such will continue to work closely with El-Sayeh et al.


G. A. Silver Bristol-Myers Squibb, Uxbridge, UK
R. Gutierrez-Esteinou Bristol-Myers Squibb, Princeton, New Jersey, USA, Email: rolando.gutierrez-esteinou@bms.com
R. D. McQuade Otsuka America Pharmaceutical, Princeton, New Jersey, USA, Email: doi: 10.1192/bjp.191.1.86

Declaration of interest

The main finding of El-Sayeh et al's systematic review that ‘compared with placebo, aripiprazole treatment was associated with a significant decrease in relapse rates, increased compliance with the study protocol, and a decrease in prolactin levels below the expected values’ is overshadowed by a background of complaint about lack of data. What one wants to know is what was the spectrum of activity with respect to symptoms? A substantial body of data was collected with standard rating scales on 4125 patients in ten separate trials. With no single exception, El-Sayeh et al record that these data were either ‘unusable’ or that standard deviations were not available (Table 1); they therefore conducted no analysis.

It seems incredible that after contacting relevant authors and the manufacturers of aripiprazole El-Sayeh et al came up with such a barren yield. There surely are data available and a systematic reviewer has a duty to obtain them and make them available in comparative form.

A more serious deficiency relates to the authors’ clear innuendo that reports of deaths which are possibly drug related have not been widely disseminated. They further argue that ‘not disseminating clear information regarding these people’s outcomes... breaks that unspoken contract that occurs between researchers and trial participants at the point of gaining informed consent’. In a poster presentation at the Winter Workshop on Schizophrenia Research in February 2006 the authors were even more explicit, ‘In two studies 8 people allocated aripiprazole died. Even if the mortality of people with schizophrenia is 2–3 times that of the general population, the age-standardised death rate in these studies exceeds even that pessimistic estimate by 400–500 percent... Mortality data are concerning’. To make the point crystal clear the poster included a representation of a coffin.

It appears that El-Sayeh et al made a simple mistake – they thought that a number of deaths recorded in trials on the Food and Drug Administration (FDA) website (http://www.fda.gov) related differentially to patients on aripiprazole, whereas in fact these deaths were in the uncontrolled follow-up phase and were neither selective to aripiprazole, nor in excess relative to age and gender norms. The data were accessible, and were known to the FDA and to the relevant companies. Authors of systematic reviews no doubt have a duty to draw attention to deficiencies of trial data as they see them, but they also have a
responsibility to marshal all the findings in a scientifically revealing way. If they make an error they have a duty to correct it.

Declaration of interest
T.J.C. is a co-organiser of the Winter Workshop for Schizophrenia Research which has been supported by Bristol-Myers Squibb and Otsuka, Janssen Pharmaceuticals, Lilly, Lundbeck and Astra Zeneca.


T. J. Crow SANE Prince of Wales Centre, Warneford Hospital, Oxford OX3 7JX, UK. Email: tim.crow@psych.ox.ac.uk doi: 10.1192/bjp.191.1.86a

Author’s reply: We generally concur with the views of Silver et al and we continue to collaborate with those at Bristol-Myers Squibb and Otsuka to obtain data which were not easily accessible to us when the review was initially conducted. The updated version of this review is much improved by the incorporation of these data (El-Sayeh & Morganti, 2006). The original version, however, was submitted to the Journal in June 2004.

We were interested that the review fell short of Professor Crow’s expectations. Perhaps he is correct in saying that there is a grumbling background to the whole review but it was peer reviewed and there was no objection to this. Professor Crow was surprised that our searches came up with such a ‘barren yield’ of data. Perhaps his experience in this area is not ours. We asked employees of Bristol-Myers Squibb to check our review. Those that kindly visited us and promised data are explicitly mentioned in widely accessible versions of this review (El-Sayeh & Morganti, 2004). Other authors referred us to the company for additional information. Professor Crow goes on to say that it is the duty of systematic reviewers to make data available in comparative form. We have tried to be fair, open and explicit with what we could get. If Professor Crow can get more data we will of course be grateful for those.

Professor Crow draws attention to aripiprazole and mortality as presented in a poster at the Winter Workshop on Schizophrenia Research in February 2006. After publication of our paper in the Journal we obtained clarification from Bristol-Myers Squibb regarding the eight deaths. This information was forthcoming precisely because of the poster presentation in 2006. Two years earlier we had met with representatives of the company and asked for clarification of our results before publication in the Cochrane Database of Systematic Reviews and a note of this meeting is made (El-Sayeh & Morganti, 2004). The offer of clarification and further contact did not materialise until after the poster presentation. Thereafter Bristol-Myers Squibb showed us how we had indeed failed to note how these people had died in the post-randomisation protracted follow-up of the two studies in question, so normalising the seemingly alarming standardised mortality ratio previously presented (El-Sayeh & Morganti, 2004). We do not think anyone would say that these data were easy to locate or are clear (Dubitsky et al, 2002), although Professor Crow may think otherwise.

As mentioned in our review, currently available data do not seem to support the prolific use of aripiprazole. In suggesting otherwise, there may be a danger of giving false hope to clinicians and recipients of care.


H. G. El-Sayeh Briary Wing, Harrogate District Hospital, Harrogate HG2 7JX, UK. Email: hany.el-sayeh@chrd-pct.nhs.uk

C. Morganti Ospedale Niguarda Ca’Granda, Milan, Italy

C. E. Adams Academic Unit of Psychiatry, University of Leeds, Leeds, UK doi: 10.1192/bjp.191.1.87a

One hundred years ago

John Murray’s Royal Asylum

The statistics of insanity are perhaps more lacking in precision of terms than are those relating to any other human affairs. Chief among the elastically uncertain stands the term “recovery”. Yet on it depends the true history both in the positive and negative sense of our fight with the disease. Dr. Urquhart gives his interpretation of the term, and we consider that it is as fair and accurate as can be looked for:

“The number of readmissions (15) was disproportionate. The number of those suffering from recurrence of mental disorder (22) was also disproportionate.

In these observations the word ‘recovery’ is used to indicate those in whom there is re-establishment of mental soundness permitting of the return of the patient to his place in the world without requiring the care and supervision of others. The lucid interval may prove to be of life-long duration, it may last for years, or only for months. Doubts have been expressed regarding the propriety of liberty in many of these cases. It has been represented as a wrong to the lies. This is a new phase of opinion. For many years we have been accustomed to accusations of undue detention in asylums, elaborate safeguards have been devised to protect the insane from that evil, and now the tide of opinion seems to be setting in the contrary direction. As the law stands there is no longer authority for the detention of a person after he ceases to be insane; and, in the great majority of cases, it would be an intolerable hardship to be detained indefinitly because of a possibility of untoward remote consequences. No doubt there are those, including many who have never been under custodial care, who should be limited in liberty of action under revised legal enactments; but the advocates of extreme measures will have to be content with less Spartan remedies than they formulate. The practice of discharge on recovery, or even on improvement, may entail occasional hardships, but on the whole it is appropriate to existing conditions.

REFERENCE

Book reviews

EDITED BY SIDNEY CROWN, FEMI OYEBODE and ROSALIND RAMSAY

Trauma, Truth and Reconciliation: Healing Damaged Relationships

This is a philosophical look at a subject of intense interest to psychiatrists. The thirteen authors include seven philosophers, two theologians, three psychologists and six psychiatrists, several doubly qualified. Some chapters are unashamedly theoretical; use of conditional tenses is quite refreshing as a change from medical books.

Potter, the editor, introduces the topic of the healing power of forgiveness and reconciliation, including the vexed issue of the morality of forgiving someone with no remorse. Brendel, considering the implications of the Truth and Reconciliation Commission (TRC) for psychotherapy, looks at the way that apology can be an important first step towards repairing a damaged relationship. Kruger writes about forgiveness and reconciliation from a South African perspective. She regards Desmond Tutu’s ubuntu theology as having been a guiding force in the work of the TRC. Zachar, also considering the TRC from a psychotherapeutic perspective, argues that it had a legitimate purpose in the management of both individual and societal rage, and that the experience of rage is related to a desire for justice. He deals with the different problem of prescribed forgiveness. How the philosophical approach to the concept of truth helps us understand the notion and role of truth-telling in our lives is linked by Mitchell to the work of the TRC.

Political reconciliation, the process of building healthier relationships among citizens formerly estranged through conflict or repression, is discussed by Murphy. Psychotherapy can help to clarify past events and explore feelings but cannot provide reconciliation, according to Spitz, who elaborates how truth and trust are required. Rawlinson discusses the moral significance of the act of forgiveness. The hypothesis of Glas is that the description of the dynamic of evil helps us to understand forgiveness and reconciliation, and that this is useful for the work of the psychiatrist. ‘It is characteristic for forgiveness to have insufficient grounds; this is not a weakness but indicates forgiveness’s power . . . .’ Verhagen develops a relational model of forgiveness involving both victim and offender.

How forgiveness may reinforce gender stereotypes and uneven power relations is discussed by Lamb. Perring states that truth-telling is an essential ingredient in the healing process and discusses consumer/survivor movements in psychiatric care. ‘Change the story, the future changes’ is central to work with aboriginal people in Canada, according to Mehl-Madrona who cooperates with community elders in psychotherapy.

This book is recommended to those who want to take one of the fundamentals of our work a bit further.

Andrew Sims c/o Royal College of Psychiatrists, 17 Belgrave Square, London SW1X 8PG, UK. Email: ruth-andrewsims@ukgateway.net doi: 10.1192/bjp.191.1.88

Asylum to Action: Paddington Day Hospital, Therapeutic Communities and Beyond

I was pleasantly surprised and excited when asked to review this book. Initially apprehensive, I found the experience enriching, not only in expanding my limited knowledge about therapeutic communities but also in awakening a sense of intrigue and curiosity.

Asylum to Action provides a glimpse into history, of the rise and subsequent demise of the Paddington, a therapeutic community set in London in the late 60s/early 70s. The author, Helen Spandler, is a Research Fellow in social work and has published in other areas of mental health. Asylum to Action is a summary of her retrospective work on Paddington. She is clearly very knowledgeable and passionate about the research she carried out into the history of the hospital. This is conveyed within her narrative and she achieves a good balance between giving factual information and providing a forum in which to pose questions and allow productive debate.

The book is divided into thoughtful chapters. It begins by giving a history of the formation of therapeutic communities in an era of anti-psychiatry, social ideals and a rise in patients with mental health problems using their own voices to promote change.

The Paddington appears to have been an idealised experiment into health democracy and libertarianism, at a time of immense political struggle within the healthcare system. When faced with uncertainty in 1972, a tremendous effort by the Paddington’s staff and patients prevented its closure. It provided a foundation from which patients gained more rights in treatment choices and allowed for an escape from traditional medicalisation and labelling. Unfortunately, it also portrays a very real situation where power struggles, a lack of boundaries and regulations and the extremes of political correctness can lead to corruption and destruction.

In contrast to other literature on the Paddington Day Hospital, Spandler challenges the negative accounts. Rather than purely lingering on its downfall, Spandler reinforces its achievements as a therapeutic microcosm that provided benefits to its
patients. Its success also infiltrated the wider social context of the media and general public.

I believe that one of the rewarding aspects of this book is that aside from its core content it covers aspects of the ongoing struggles of the National Health service in today’s political climate and the importance of solidarity.

Having spent some time in a therapeutic community in Massachusetts, USA as a medical student on my elective, I was drawn by the parallels of my own experiences. I became quite involved in the various dilemmas and debates that the book threw into question.

This book is a light and easy read. Although it may not be seen as a core text in terms of psychiatric training programmes, I would recommend it as an interesting and controversial read, for both mental health professionals and a wider audience.

Marianna Leontis
Enfield Town Team CMHT,
59-60 Silver Street, Enfield EN1 3EP, UK.
Email: marianna.leontis@hotmail.com
doi: 10.1192/bjp.191.1.88a

An Introduction to the Psychotherapies (4th edn)
ISBN 0198520921

Although this book is nearly 30 years old, I had never seen it before. First impressions are favourable. A multi-author guide to the range of current psychotherapies through 18 chapters, including ones on psychotherapy with young people and the elderly, it nearly always succeeds in delivering on the essentials. These are: the key principles of an approach; when it is likely to be used; and the extent to which there is evidence for its effectiveness. A sense of consideration for the reader’s likely needs informs the book so thoroughly that it is surprisingly readable, with several chapters likely to become recommended introductions for a given approach. The book’s value is augmented by summary lists of key readings, where most contributors are commendably objective in their selections.

A brief historical introduction allows some reference to be made to areas, such as person-centred psychotherapy, which are thereafter effectively ignored. While no book of this kind is likely to be completely comprehensive, three omissions were noticeable given the likely needs of trainee psychiatrists. First, although several important models of brief psychotherapy are considered in some depth (based on the work of Malan & Davanloo; Ryle & Hobson), the one that is now the most widely used, Klerman’s interpersonal psychotherapy, is not. Second, supportive psychotherapy (the darling of many MRCPsych examination essays) retains a chapter, but in it, as elsewhere, the newcomer to the field is given no guidance on what the term counselling means, or on how it might differ from psychotherapy. Third, one apparent consequence of the prioritisation of general principles here is a failure to illustrate psychotherapy when practised other than in out-patient settings. (The chapters on behavioural therapy and psychotherapy for older people are exceptions here, but one on ‘family therapy in the adult psychiatric setting’ is misleadingly titled. Moreover, there is no acknowledgment at all of the role of specialist hospitals or therapeutic communities.) These caveats aside, I recommend trainee psychiatrists have access to a copy of this reliable compendium throughout the early years of training.

Chris Mace
St Michael’s Hospital, Warwick
CV34 5QW, UK. Email: c.mace@warwick.ac.uk
doi: 10.1192/bjp.191.1.89
From the Editor’s desk

PETER TYRER

ANTICIPATING THE FUTURE

I was recently in the University of Uppsala, the oldest university in Sweden, which this year celebrates the 300th anniversary of its greatest scientist, Carl Linnaeus – although he has had to jostle with other notables such as Carl Wilhelm Scheele, the discoverer of oxygen in 1772, who did not have high impact factor publications to disseminate his work and was beaten into second place by Joseph Priestley in 1774, who did, and Anders Celsius, who devised the centigrade scale of temperature in 1742. All of them were great anticipators of the future, and the impossible task of the editor of a learned journal is to separate those contributions that are ephemeral or misleading from the pure gold of discovery that has lasting impact.

So which of this month’s articles may appear humdrum or trivial today but inestimably important in 100 years? Will it be Chiara Samele and her colleagues (pp. 1–2), who anticipate the Great Consumer Hope that patients will determine their own treatment in the future? Or will it be Davies et al’s (pp. 14–22) dramatic demonstration that old is best when it comes to choosing cost-effective antipsychotic drugs? And once chosen, will we have the universal spread of vocational centres that drugs? And once chosen, will we have the universal spread of vocational centres that really knows, but the impressive results of translation on the DISC-1 gene, shown to be related to psychosis in a Scottish family (Millar et al, 2001; Muir et al, 2006), to the mouse in the laboratory. It has been argued that such linkage studies are completely misleading (Crow, 2007) and what is clearly needed is a better model of the schizophrenic mouse for researchers. As someone who is at the opposite pole of translational research than Low and Hardy I have the following tips to help in identifying this pathological animal through clinical assessment.

Primary delusion (belief that the mouse is carrying out an experiment on the investigator): mouse noticed to be observing the investigator more closely than the investigator does the mouse.

Knight’s move thinking: mouse observed to move two steps in one direction and one to the left or right repeatedly.

Visual hallucinations and paranoid delusions: mouse believes tail to be a snake and attempts to run away from it; when faced by several other mice believes their tails represent a threatening conspiracy and may attack them (heads you lose but tails you win).

Delusional mood: mouse and investigator both feel there is something funny going on but cannot quite figure out what it is.

Passivity: mouse believes its mind is under the total control of the investigator and is observed waiting for instructions.

A copy of the new edition of Frank Fish’s book (Casey & Kelly, 2007) is a useful guide for further signs of this fascinating disorder.


